U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-293 615

Methods Development for Assessing Air Pollution Control Benefits. Volume I. Experiments in the Economics of Air Pollution Epidemiology

Wyoming Univ, Laramie

Prepared for

Environmental Protection Agency, Washington, DC

Feb 79

United States
Environmental Protection
Agency

Office of Health and
Ecological Effects
Washington DC 20460

EPA-600/5-79-001a
February 1979

Research and Development

PB 293615

EPA

Methods Development for Assessing Air Pollution Control Benefits

Volume I, Experiments in the Economics of Air Pollution Epidemiology

MAR 28 1979

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

OTHER VOLUMES OF THIS STUDY

Volume II, Experiments in Valuing Non-Market Goods: A Case Study of Alternative Benefit Measures of Air Pollution Control in the South Coast Air Basin of Southern California, EPA-600/5-79-001b.

This volume includes the empirical results obtained from two experiments to measure the health and aesthetic benefits of air pollution control in the South Coast Air Basin of Southern California.

Volume III, A Preliminary Assessment of Air Pollution Damages for Selected Crops within Southern California, EPA-600/5-79-001c.

This volume investigates the economic benefits that would accrue from reductions in oxidant/ozone air pollution-induced damages to 14 annual vegetable and field crops in southern California.

Volume IV, Studies on Partial Equilibrium Approaches to Valuation of Environmental Amenities, EPA-600/5-79-001d.

The research detailed in this volume explores various facets of the two central project objectives that have not been given adequate attention in the previous volumes.

Volume V, Executive Summary, EPA-600/5-79-001e.

This volume provides a 23 page summary of the findings of the first four volumes of the study.

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/5-79-001a	2.	3. RECIP PB293615
4. TITLE AND SUBTITLE Methods Development for Assessing Air Pollution Control Benefits: Volume I, Experiments in the Economics of Air Pollution Epidemiology		5. REPORT DATE February 1979 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Thomas D. Crocker, William D. Schulze, Shaul Ben-David, and Allen V. Kneese		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Wyoming Laramie, Wyoming 82071		10. PROGRAM ELEMENT NO. 1HA616 and 630 11. CONTRACT/GRANT NO. R805059-01
12. SPONSORING AGENCY NAME AND ADDRESS Office of Health and Ecological Effects Office of Research and Development U.S. Environmental Protection Agency Washington, DC 20460		13. TYPE OF REPORT AND PERIOD COVERED Interim Final, 10/76-10/78 14. SPONSORING AGENCY CODE EPA-600/18
15. SUPPLEMENTARY NOTES		

16. ABSTRACT

This volume employs the analytical and empirical methods of economics to develop hypotheses on disease etiologies and to value labor productivity and consumer losses due to air pollution-induced mortality and morbidity.

In the mortality work, 1970 city-wide mortality rates for major disease catagories have been statistically associated with aggregate data from sixty U.S. cities on physicians per captia, per capita cigarette consumption, dietary habits, air pollution and other factors. The estimated effect of air pollution on mortality rates is about an order of magnitude lower than some other estimates. Nevertheless, rather small but important associations are found between pneumonia and bronchitis and particulates in air and between early infant disease and sulfur dioxide air pollution.

The morbidity work employed data on the generalized health states and the time and budget allocations of a nationwide sample of individual heads of household. For the bulk of the dose-response expressions estimated, air pollution appears to be significantly associated with increased time being spent acutely or chronically ill. Air pollution, in addition, appears to influence labor productivity, where the reduction in productivity is measured by the earnings lost due to reductions in work-time.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Economic analysis Air pollution Epidemiology Public health	Economic benefits of pollution control	13B

18. DISTRIBUTION STATEMENT Release unlimited	19. SECURITY CLASS *(This Report)* Unclassified	21. NO. OF PAGES 177
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE PCA09 / MF A01

EPA Form 2220-1 (9-73)

☆U.S. GOVERNMENT PRINTING OFFICE:1979 620-007/3769 1-3

ERRATA

Please make the change shown below in

METHODS DEVELOPMENT FOR ASSESSING AIR POLLUTION CONTROL BENEFITS,

Volume I, EPA-600/5-79-001a, February 1979, p. 154

Volume V, EPA-600/5-79-001e, February 1979, p. 8

Change $0.071 to 0.071 cents in the first arithmetic equation on these two pages.

EPA-600/5-79-001a
February 1979

METHODS DEVELOPMENT FOR ASSESSING AIR POLLUTION CONTROL BENEFITS

Volume I

Experiments in the Economics of Air Pollution Epidemiology

by

Thomas D. Crocker and William Schulze
University of Wyoming
Laramie, Wyoming 82071

Shaul Ben-David
University of New Mexico
Albuquerque, New Mexico 87131

Allen V. Kneese
Resources for the Future
1755 Massachusetts Avenue, N.W.
Washington, D.C. 20036

USEPA Grant #R805059010

Project Officer
Dr. Alan Carlin
Office of Health and Ecological Effects
Office of Research and Development
U.S. Environmental Protection Agency
Washington, D.C. 20460

OFFICE OF HEALTH AND ECOLOGICAL EFFECTS
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

1a

DISCLAIMER

This report has been reviewed by the Office of Health and Ecological Effects, Office of Research and Development, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

PREFACE

The motivation for this volume originated in the authors' mutual and reinforcing convictions that economic analysis and its techniques of empirical application could contribute to the resolution of certain puzzles in studies of the incidence and severity of diseases in human populations, particulary the epidemiology of air pollution. The prior works of Lester Lave, Eugene Seskin, and V. Kerry Smith have provided an excellent base from which to initiate our efforts. These researchers, in addition to Dennis Aigner, Shelby Gerking, Leon Hurwitz, and Roland Phillips have also provided many worthwhile comments and criticisms. None of these individuals are responsible, however, for the results we have obtained.

ABSTRACT

This study employs the analytical and empirical methods of economics to develop hypotheses on disease etiologies and to value labor productivity and consumer losses due to air pollution-induced mortality and morbidity. Since the major focus is on methodological development and experimentation, all the reported empirical results are to be regarded as tentative and ongoing rather than definitive and final.

Two experiements have been conducted. First, using aggregate data from sixty U.S. cities, 1970 city-wide mortality rates for major disease categories have been statistically associated with aggregate population characteristics such as physicians per capita, per capita cigarette consumption, dietary habits, air pollution and other factors. Dietary variables, smoking, and physicians per capita are highly significant statistically. However, the estimated contribution the latter variable makes to reducing mortality rates becomes evident only after we recognize that human beings attempt to adjust to disease by seeking out more medical care. The estimated effect of air pollution on mortality rates is about an order of magnitude lower than some other estimates. Nevertheless, rather small but important associations are found between pneumonia and bronchitis and particulates in air and between early infant disease and sulfur dioxide air pollution.

The second experiment, which focused on morbidity, employed data on the generalized health states and the time and budget allocations of a nationwide sample of individual heads of household. For the bulk of the dose-response expressions estimated, air pollution appears to be significantly associated with increased time being spent acutely or chronically ill. Air pollution, in addition, appears to influence labor productivity, where the reduction in productivity is measured by the earnings lost due to reductions in worktime. The reduction in productivity and to air pollution-induced chronic illness seems to be much larger than any reductions due to air pollution-induced acute illness.

CONTENTS

Page

Chapter I: Introduction to Volume I 1

Chapter II: Some Issues 3
2.1 Epidemiology and Economics 3
2.2 When Microeconomics Doesn't Matter 5
2.3 When Microeconomics Does Matter 9
2.4 The Costs of Pollution-Induced Disease 12

Chapter III: Sources of Error 15
3.1 Problems in Statistical Analysis 15
3.2 Heteroskedasticity 15
3.3 Multicollinearity 16
3.4 Causality and Hypothesis Testing 17
3.5 Aggregation 19

Chapter IV: The Sixty-City Experiment 24
4.1 Objectives of the Experiment 24
4.2 Value of Life Vs. Value of Safety 27
4.3 A Methodological Basis: Does Economics Matter? 32
4.4 The Sixty-City Data Set: Selection of Variables 35
4.5 Empirical Analysis 53
4.6 A Tentative Estimate of the Value of Safety from Air Pollution Control 70

Chapter V: The Michigan Survey Experiment 72
5.1 Objectives of the Experiment 72
5.2 The Sample and the Variables 72
5.3 Estimates of Dose-Response Rates for Acute and Chronic Illness . . . 96
5.4 Recursive Estimates of the Effect of Air Pollution Upon Health, Labor Earnings, and Hours of Work 119
5.5 A Model of the Effect of Air Pollution on the Demand for Health 137
5.6 Some Empirical Results: The Demand for Freedom from Air Pollution-Induced Acute and Chronic Illness 142
5.7 Overview of Empirical Results 148

CONTENTS
(continued)

Page

Chapter VI: An Estimate of National Losses in Labor Productivity Due to Air Pollution-Induced Morbidity 153
6.1 Introduction 153
6.2 The Assumptions 154

FIGURES

Number Page

2.1 Alternative Measures of Disease Incidence 7

3.1 Marginal Purchase Price and Marginal Willingness-to-Pay . . . 22

4.1 Hypothetical Human Dose-Response Function 26

6.1 A Representation of the Effect of Air Pollution Upon Labor Productivity 153

TABLES

Number Page

4.1 Objectives and limitations 28

4.2 Mortality Variables 36

4.3 Dietary Variables 37

4.4 Social, Economic, Geographic, and Smoking Variables 38

4.5 Air Pollution Variables 39

4.6 Sources of Data 41

4.7 Simple Correlation Matrix for 1965 Diet Variables 43

4.8 Simple Correlation Matrix for Air Quality Variables 45

4.9 Simple Correlation Matrix for Included Variables 46

4.10 Complete Listing of Data Used 47

4.11 Summary of Two-Stage Linear Estimates of Factors in Human Mortality Hypotheses not Rejected at the 97.5% Confidence Level (One-tailed t-test, $t \geq 2.0$) 54

4.12 Reduced Form Equation 56

4.13 Total Mortality 57

4.14 Vascular Disease 58

4.15 Heart Disease 59

4.16 Pneumonia and Influenze 60

4.17 Emphysema and Bronchitis 61

4.18 Cirrhoris 62

4.19 Kidney Disease 63

4.20 Congenital Birth Defects 64

TABLES
(continued)

Number Page

4.21 Early Infant Diseases 65

4.22 Cancer 66

4.23 Methodology for Health Benefits Assessment 69

4.24 Urban Benefits from Reduced Mortality: Value of Safety for 60% Air Pollution Control 71

5.1 Complete Variable Definitions 74

5.2a Representative Means and Standard Deviations of Health and Air Pollution Variables for Samples Involving Family Heads Currently Employed or Actively Looking for Work* 78

5.2b Representative Means and Standard Deviations of All Other Variables[a] 79

5.3 Expected Signs for Explanatory Variables in Estimated Dose-Response Functions 88

5.4 Proportions of Entire Survey Research Center Sample Processing a Particular Characteristic During 1971 93

5.5 Matrix of Simple Correlation Coefficients for a 1971 Representative Dose-Response Function Sample 99

5.6a Dose-Response Rates for ACUT: Unpartitioned Samples 100

5.6b Dose-Response Rates for ACUT: Partitioned Samples 104

5.7a Dose-Response Rates for LDSA: Unpartitioned Samples 105

5.7b Dose-Response Rates for LDSA: Partitioned Samples 109

5.8 Lagged Effects of Total Suspended Particulates Upon Duration of Chronic Illnesses (LDSA) of Respondents Who, as of 1975, Had Always Lived in the Same State 117

5.9 Simple Correlation Coefficients Between Labor Supply and Certain Other Variables for a 1970 Sample 120

5.10a Empirical Results for a 1971 Sample Recursive Labor Supply 124

5.10b Empirical Results for a 1970 Sample Recursive Labor Supply . 126

5.10c Empirical Results for a 1971 Sample Recursive Labor Supply . 127

TABLES
(continued)

Number | Page

5.10d Empirical Results for a 1969 Sample Recursive Labor Supply . 128

5.11a Labor Supply Effects of Air Pollution-Induced Chronic and/or Acute Illnesses 129

5.11b Value of Labor Supply Effects of Air Pollution-Induced Chronic and/or Acute Illnesses for Pollution Concentrations Two Standard Deviations Removed from the Mean Concentration . 131

5.12 Willingness to Pay to Avoid Acute Illness 145

5.13a Two-Stage Least Squares Estimates of WAGE Expressions for Chronic Illness 146

5.13b Two-Stage Least Squares Estimates of Chronic Illness Expressions (LDSA) 147

6.1 Major Assumptions Limiting Generality of Results 155

6.2 Distinguishing Features that Enhance the Generality of Results 157

6.3 Estimated Expressions to be Used to Calculate the Effect of Air Pollution-Induced Illness on Labor Productivity 158

6.4 Estimated Per Capita Aggregate Gains in 1970 U.S. Labor Productivity Due to a 60 Percent Reduction in Air Pollution . 161

CHAPTER I

INTRODUCTION TO VOLUME I

Volume I focuses on developing methodology for valuing the benefits to human health associated with air pollution control. Air pollution may affect human health in three ways: (1) by increasing mortality rates, (2) by increasing the incidence and the severity of chronic illness (morbidity), and (3) by increasing the incidence and the severity of acute illness (morbidity).

A number of approaches for determining health effects and valuing them in economic terms are developed within the study. First, if a dose-response relationship is known between mortality rates and air pollution or between days lost from work due to illness (productivity loss) and air pollution, economic losses can be approximated. In the former case, one must know how consumers value increased safety. Thus, if air pollution control reduces risk of death from air pollution related disease, studies of the value consumers place on safety in other situations -- on the job, in transportation, etc. -- can be applied to measuring the benefits of pollution control programs. Note, however, that valuing safety for small changes in risk is very different from the alternative of valuing human life through lost earnings -- an approach rejected here. Rather, the focus is on examining the value of safety to individuals; that is, how much consumers are willing to pay for safety obtained through pollution control. For morbidity losses, lost time from work and lost productivity during hours of work can be relatively easily valued using observed wage rates.

A second approach for valuing the effects of air pollution on human health is to attempt to observe the effect of air pollution directly on economic factors, thus avoiding the necessity of developing dose-response relationships. If one can develop relationships employing data on wages, wealth, socioeconomic and health status characteristics as well as air pollution concentrations, consumer willingness to pay to avoid illness can be derived. We term this second methodology the willingness to pay approach. It is based on traditional microeconomic theory.

Volume I contains two experiments. First, a data set on sixty U.S. cities is explored to determine if some of the problems of aggregate epidemiology -- epidemiology using aggregate data on groups of individuals as opposed to data on individuals -- can be overcome. The study attempts to estimate a human dose-response function wherein city-wide mortality rates for major disease categories in 1970 are statistically related to population characteristics such as doctors per capita, cigarettes per capita, information on dietary patterns, race, age and air pollution. The study is unusual in two respects. First, it is the first such aggregate epidemiological study of the effect of air pollution on mortality to include dietary variables, which, along with smoking and medical care, prove to be highly significant. Second, it may be the first study using aggregate data to account for the fact that human beings will attempt to adjust to disease by seeking out more medical care. Thus, cities with high mortality rates are likely to have more doctors per capita. This adjustment process has in the past prevented an estimate of the direct effect of doctors on the prevention of disease. An estimation technique for handling this bias problem is employed, which identifies the contribution medical care makes in reducing mortality rates. The impact of including these new variables in the analysis is substantial.

The second experiment focuses on morbidity rather than mortality. It employs data on the health and the time and budget allocations of a random sampling of the civilian population nationwide. The sample, which was collected by the Survey Research Center of the University of Michigan, consisted of approximately 5,000 heads of households for nine years from 1967 through 1975. Generalized measures of acute illness, stated in terms of annual work-days ill, and of chronic illness, stated in terms of years ill, are available.

The procedures used to estimate dose-response expressions have two somewhat unusual features: (1) care has been taken to employ as explanatory variables only those factors not influenced by the individual's current decisions or health status; and (2) by randomly drawing different samples of individuals, substantial effort was devoted to replicating results.

This volume begins in Chapter II by discussing the role of economic analysis in epidemiology. We then introduce in Chapter III the formidable list of statistical problems faced by epidemiological analysis of air pollution. Finally, Chapters IV and V present the Sixty-City and Michigan Survey Experiments, respectively. Chapter VI presents additional economic results on the valuation of air pollution-induced morbidity.

Chapter II

SOME ISSUES

2.1 Epidemiology and Economics

The motivation for this volume originated in the authors' mutual and reinforcing convictions that economic analysis and its techniques of empirical application could contribute to the resolution of certain puzzles in studies of the incidence and severity of diseases in human populations, particularly the epidemiology of air pollution. The results of our initial efforts to provide empirical support for this perspective are presented in succeeding chapters. Before proceeding to these chapters, however, it is necessary, in order to display the basic rationale for our empirical efforts, to explain our position that economics has some worthwhile things to offer epidemiology.

Many reviews of the epidemiological literature dealing with pollution have remarked upon the relative lack of consistent findings across studies for the effects thought to be caused by any one pollutant. Various reasons are typically advanced for these inconsistencies: inadequate characterization of the pollutants; the use of noncomparable, and sometimes questionable, estimating techniques; failure to account for other environmental influences and self-induced health stresses such as ambient temperature and cigarette smoking; failure to distinguish between pollution levels at work and at home; insufficient attention to differences in genetic endowments, and other factors. The list is sufficiently long and repetitive to be reminiscent of the beat of a somber military cortege. This march has two elements: measurement error and specification error.

The first error element refers to the fact that some variables included in epidemiological studies are inaccurately measured. Sources of error of this sort, however, are hardly unique to epidemiology. They are at least equally common in empirical applications of economic analysis and will therefore be accorded our scrutiny when we discuss our empirical efforts. For the moment, we wish to consider those possible sources of specification error in epidemiological studies that have a basis in the microeconomic theory of the behavior of the individual human being. Our fundamental point is that human beings, the objects of epidemiological attention, make purposive choices with respect to health states and phenomena that influnce health states. To the extent that health states are a result of the individual's purposive acts, one must explain these acts in order to comprehend the determinants of the health state. Microeconomics provides a means for grasping the determinants of the individuals's purposive acts.

Acceptance of this perspective adds another dimension (in addition to the social provision of preventive and ameliorative medical inputs) by which social policy can influence the health states of the population, i.e., those factors that influence choices of acts affecting health states can serve as policy instruments.

Specification error occurs in epidemiology (and in economics) when some varibles relevant to the explanation of variations in the health state of interest are improperly introduced or are altogether excluded from the analysis. The biased and incosistent estimates that are the result of excluding nonorthogonal explanatory variables from an expression to be estimated are well-known and intuitively obvious in any case. One can hardly, for example, expect to obtain an accurate estimate of the impact of cigarette smoking on circulatory diseases if the ages of the sample individuals are not controlled. Less obvious, however, are the reasons why common economic variables such as prices often are relevant to epidemiological analyses and why certain variables, both biologic and economic, are sometimes improperly introduced to these analyses.

Some of the most widely known findings in the epidemiology literature concern the respiratory effects (cancer, acute bronchitis, emphysema, the common cold, and pneumonia) of air pollution. View the absence of these respiratory effects as an output that can be reduced by various combinations of clean air and ameliorative medical care, where the latter are considered to be inputs. The literature suggests that there are significant differences in the input-input ratios and in the input-output ratios among various locales, where these locales frequently differ in population size. Suppose it has been observed that:

1. Per capita absence of respiratory diseases is inversely associated with city size.

2. Per capita availability of ameliorative medical care is directly associated with city size.

3. Per capita absence of respiratory disease is directly associated with per capita availability of clean air and ameliorative medical care.

4. Per capita clean air is inversely associated with city size.

5. Respiratory disease absence per unit of clean air and ameliorative medical care is directly associated with city size.

Do the five observations have sufficient informational content to justify a judgment that the dirty air often found in large population concentrations is associated with greater incidence of respiratory diseases and is therefore a plausible cause of these diseases? It would not be surprising if different epidemiological investigators drew a variety of largely contradictory conclusions about the relationships between respiratory diseases, clean air, and ameliorative medical care from these five observations. Contradictions are perhaps inevitable because the ratios expressed in the observations will often be inappropriate means by which to attempt to make judgments about the relative susceptibilities of human beings to respiratory diseases.

An intuitive notion of the incidence of a disease refers to the frequency of occurrence, given particular levels of instigating factors. Intuition is sometimes misleading. Observation (1) suggests that small cities have less incidence because they have less respiratory disease. Observation (5) leads to the opposite conclusion since large cities have fewer respiratory diseases relative to their clean air. But observation (4) makes small cities look favorable because of their greater provision of clean air. Or do large cities subject their populations to greater incidence of respiratory effects by having fewer units of ameliorative medical care available? Observation (3) again favors small cities because of the greater per capita availability of ameliorative medical care.

One might suspect from (5), (4), and (2) that larger cities have more ameliorative medical care relative to clean air than do smaller cities. The former have dirtier air and thus try to compensate by providing additional ameliorative medical care. It is thus not surprising that the ratio of of absence of disease per unit of available medical care favors the larger cities. An alternative interpretation of (3) is that disease frequency increases with city size not only because of dirtier air but also because the price to the consumer of medical care is greater than in smaller cities. Greater prices of these services for the consumer can imply greater returns for the profession that provides these services. Greater returns attract these professionals, resulting in greater availability of their services. However, these same higher prices also mean that sufferers from a respiratory disease of given severity will seek out less ameliorative medical care. Are then these prices, the dirty air, or the consumption of medical care the causes of the incidence of the respiratory disease? Recognition that they are intertwined is a significant but insufficient step. The nature of the intertwining remains to be explained.

2.2 When Microeconomics Doesn't Matter

Microeconomic analysis specifies the conditions under which decisionmakers (human beings) are expected to have identical ratios of inputs and outputs. Basically, these identical ratios would occur if: (1) all decisionmakers had identical biological endowments and transformed inputs into health states in precisely the same fashion; (2) all decisionmakers faced the same prices in (implicit and explicit) input and health state markets; (3) all decisionmakers had the same real income; and (4) all decisionmakers had identical preference orderings. If all these conditions were fulfilled with respect to a particular pollutant, only one point could be observed on the epidemiologist's dose-response curve: there would be no variation whatsoever in the observable behavior of individuals.

We nevertheless observe decisionmakers in the real world with similar states-of-health who have different biological endowments and varying ways of transforming inputs into these health states. One can, of course, pass muster in explaining the real world by assuming that decisionmakers (?) behave randomly or that all health states, whether present or future, are determined by physical or biological factors beyond the decisionmaker's present control.

This is no different than assuming that the decisionmaker is abysmally ignorant of cause-and-effect with respect to health states or that he just does not care about his health state. If any of the conditions in this paragraph are in fact true, then current epidemiological precedures, which tend to give short shift to economic variables and which implicitly treat the individual as being completely unable to exercise influence over events that affect his choices, are entirely satisfactory. This abrupt statement requires clarification.

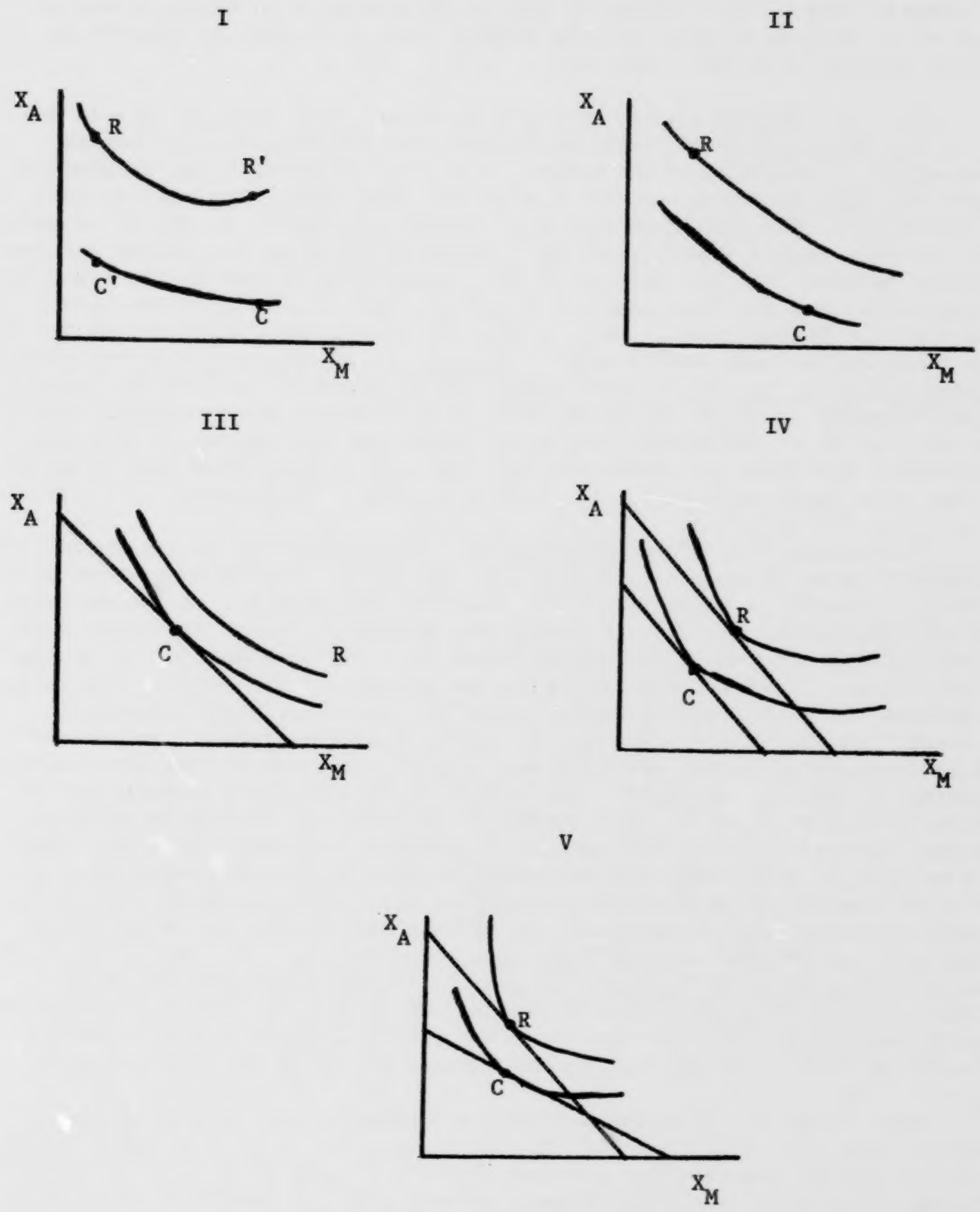
Panels I through VI of Figure 2.1 represent two unit isoquants (loci of points showing all combination of two inputs that will yield equal health states) for inputs X_M and X_A (e.g., medical care and clean air), with the current positions of decisionmakers R (a rural person) and C (a city person) indicated. Each isoquant represents the same state-of-health as the other isoquant. Note that the effectiveness of each input in providing the unit health state for each individual is assumed to decline progressively as more of one input is substituted for the other. Thus additional medical care becomes progressively less effective as the air becomes dirtier. Similarly, cleaner air becomes an increasingly poor substitute for medical care as less and less medical care becomes available.

All panels are drawn so that on the basis of his state-of-health per unit of clean air, decisionmaker C is in better shape than decisionmaker R. Conversely, decisionmaker R does better than C in terms of his health state per unit of medical care. In each panel, therefore, C uses relatively less clean air and R uses relatively less medical care to attain the unit health state. This situation is consistent with the previous five observations on the associations between city size, clean air, and ameliorative medical care.

Panels I and II refer to the case where the question of whether economic variables should be included in dose-response function analysis, and, if included, how to include them, need never arise. The clean air and medical care each individual requires to attain the unit health state are determined by physical and biological (technical) considerations alone. Purely economic considerations play no part. Nevertheless, the two panels do provide insights about cautions to exercise when attempting to establish dose-response functions by studying several individuals at one instant in calendar time. In Panel I, in the absence of knowledge about the isoquants of R and C, any attempt to establish the population dose-response function by averaging over the current positions of R and C is doomed to be a misrepresentation. The unit isoquants of Panel I belong to dose-response functions that differ not only by a constant term but which also embody entirely different responses of health states to particular cominations of medical care and clean air. The "average" or population dose-response function or isoquant established by pooling a single medical care-clean air combination from each isoquant will differ according to where each individual happens to be on his isoquant when he is observed. For example, the average of R and C' differs substantially from the average of R' and C. If and only if several medical care-clean air combinations for each individual were observed could a representative dose-response function be obtained. This would generally require that several observations over time be made of each individual.

Figure 2.1

Alternative Measures of Disease Incidence



In Panel II, several observations of each individual over time are not required because the isoquants belong to dose-response functions differing only by a constant term. This term could represent differences in biological endowments, childhood environment, previous lifestyles, and other factors with which epidemiologists traditionally deal. These same factors, however, could also explain the nonconstant difference between the isoquants of Panel I. Clearly, the current situation favors individual C in Panels I and II since he is able to attain the unit health state with smaller quantities of both medical care and clean air.

Panel III introduces the economic information of relative prices and the income that each individual has already decided to devote to health maintenance. Assume, for the moment, that each individual has decided to devote the same income and faces exactly the same prices for medical care and clean air. The result is that individual R is unable to attain or maintain the unit health state, although individual C, given his income and the relative prices, is fully able to do so. Individual R, due to his economic circumstances and his dose-response function, must settle for something less than the unit health state. Both biological and economic factors inhibit him from reaching the unit health state. Insofar as health states do not affect incomes and relative prices, this panel would appear to justify the common epidemiological practice of introducing incomes into a dose-response expression that is to be estimated. Panel IV, which has the incomes of the two individuals differing but presumes they continue to face identical relative prices, also seems to justify this practice. The justification is a mirage.

If the objective of epidemiological investigation is to ascertain the extent to which various physical and biological factors contribute to differences in the R and C-isoquants, then the introduction of income into a dose-response expression must reduce the estimated impact of inputs such as the medical care and clean air of Panel IV. The introduction of income is redundant. Income, along with relative prices and the form of the isoquants, determines the quantities of medical care and clean air each individual consumes. As the panels indicate, for given relative prices, the greater the individuals's income, the more health care and clean air he will consume, assuming he has not yet reached the unit health state. The quantities of medical care and clean air that enter the dose-response function estimate are thus partially determined by each individual's income. Thus the latter is a measure of the former and must capture part of the influence that would and should otherwise be attributed to clean air and medical care. Bluntly, epidemiological studies that include income reduce the odds that clean air will be seen as contributing to good health. The degree to which this reduction in odds is worthy of concern is dependent upon the extent to which income determines the consumption of clean air. The little evidence that is available indicates that at least within individual cities the association between income levels and cleaner air tends to be quite high.

Panel V depicts a situation where individuals R and C have nothing in common: they have different unit health isoquants, devote different income levels to health maintenance, and face different relative prices for medical care and clean air. Both individuals consume similar quantities of medical care but radically different quantities of clean air. Again, however, the epidemiologist interested solely in dose-response functions can safely neglect giving any attention to incomes and relative prices, for these serve only to determine the quantities of medical care and clean air consumed that directly determine health states. Nevertheless, this conclusion does not justify appealing to observations similar to those mentioned in the previous section as grounds for judging that clean air improves health states.

There are several alternative explanations for the ratios expressed in these observations. Different individuals may have different dose-response functions. Sometimes these differences may be captured by a constant term; at other times, the slopes of the functions may be dissimilar, invalidating attempts to ascertain population dose-response functions solely by observing each sample individual only once. Moreover, variations in individual incomes and in the relative prices of health inputs may be the cause of the observed ratios. This implies that the policymaker can influence the quantities of these health inputs consumed by doing nothing more than manipulating a limited set of purely economic variables. Under the conditions specified in this section, however, these variables have no bearing on estimating, via standard epidemiological procedures, the responses of the human organism to variations in the quantity of clean air.

2.3 When Microeconomics Does Matter

The preceding section employed stated, but not very visible, assumptions to arrive at the conclusion that epidemiological studies err when they devote attention to economic variables in attempting to establish dose-response functions. In particular, it was assumed that the individual had already decided the resources he would dedicate to health maintenance and that this decision did not influence any other decisions he might make. If either or both of these assumptions are inaccurate descriptions of reality, then microeconomics does matter in the determination of dose-response functions. The assumptions had the effect of removing the purposive nature of the human being from consideration: all the individual's choices were presumed to have already been made.

In implicit form, a good approximation of the expressions that epidemiologists frequently use to estimate the response of a particular mortality or morbidity effect to a particular environmental exposure is:

$$\pi_i = f_i(X,Y,Z,E,\varepsilon), \tag{2.1}$$

where π_i is the probability of the ith individual dying or becoming ill from the exposure; X is a vector of available ameliorative medical care inputs; Y is a vector indicating the individuals's socioeconomic class, medical history, ethnic group, etc.; Z is a vector of the individual's activities representing lifestyle habits such as diet and exercise regimens; E is a vector of environmental exposures that, a priori, are thought to be physical or biological instigators of the health effect; and ε is a stochastic error.

The form of $f_i(\cdot)$ is typically unknown and must therefore be approximated, perhaps by a linear expression. The coefficient attached to the exposure of interest would, given an acceptable level of statistical significance, then be interpreted as the increase in the health effect incidence caused by a one-unit change in the exposure. Would it then be reasonable to infer a dose-response association from the coefficient of the exposure variable?

The aforementioned inference would be correct if and only if it is possible to alter the environmental exposure without altering the value of any other explanatory variables in the expression. It is easy to show that this cannot be done when the structure is presumed to consist of no more than one relationship. The reason is that (2.1) contains at least two variables, the current and future levels of which are subject to at least some control by the individual; that is, during the period in which it is thought the health effect can occur, the individual can influence by his voluntary choices the magnitude of explanatory variables supposed to determine the health effect. For example, the probability of the individual suffering the health effect, π, is dependent upon the extent to which he chooses to use the available medical care and the mix and magnitude of activities he chooses to undertake. In order to explain the health effect outcome, one must also explain the structure underlying these choices. The following simple example shows one way in which π and Y, interpreted as income, might be jointly determined.

If both the π and Y functions can be linearly approximated, they can be written as:

$$\pi_i = \alpha_1 + \alpha_2 E + \alpha_3 X + \alpha_4 Y + \alpha_5 Z + \varepsilon_1 \qquad (2.2)$$

$$Y = \beta_1 + \beta_2 \pi_i + \beta_3 A + \beta_4 S + \beta_5 L + \varepsilon_2 \qquad (2.3)$$

Expression (2.2) states that the question of whether or not the individual is suffering from chronic bronchitis is related respectively to the non-cigarette bronchitis-causing agents (e.g., air pollution) to which he is exposed, the ameliorative medical care he consumes, his income, and the number of cigarettes he smokes. In turn, (2.3) states that the individual's income is determined respectively by whether or not he has bronchitis, his absenteeism rate, his schooling, and the length of time he has been on the job.

Solving (2.2) and (2.3) for π_i alone, we have:

$$\pi_i = \frac{\alpha_1 + \alpha_4 \beta_1}{1 - \alpha_4 \beta_2} + \frac{\alpha_2}{1 - \alpha_4 \beta_2} E + \frac{\alpha_3}{1 - \alpha_4 \beta_2} X + \frac{\alpha_4 \beta_3}{1 - \alpha_4 \beta_2} A + \ldots$$

$$+ \frac{\alpha_5}{1 - \alpha_4 \beta_2} Z + \frac{\alpha_4 \varepsilon_2 + \varepsilon_1}{1 - \alpha_4 \beta_2} \qquad (2.4)$$

Consider the coefficient attached to E in (2.4). If E is air pollution, (2.4) shows that an estimate of (2.2) will not yield the response of bronchitis incidence to dosages of air pollution, even though, in the language of epidemiologists, the dose-response is "adjusted" for medical care, life-style, and socioeconomic class. Instead, the coefficient for E in (2.2) will be a mix of effects due to air pollution, income, and the effect of bronchitis on income. The product of the coefficients for the latter two effects would have to approach zero in order for the response of bronchitis to air pollution alone to be obtained. For this to occur, chronic bronchitis could have no effect on the individual's income and this income could have no effect on his chronic bronchitis. Both assertions, particularly the first, are quite implausible. In fact, in the absence of further information, the sign that would be obtained for the coefficient of E in (2.2) is ambiguous since $\alpha_2 \geq 0$, $\alpha_4 \leq 0$, and $\beta_2 \geq 0$. It is entirely conceivable, if one were to estimate (2.2) alone, that one would find air pollution reducing chronic bronchitis! In any case, because the product of α_4 and β_2 is negative in sign, the effect of air pollution on health will be underestimated. One could readily obtain a similar result for Z, cigarette smoking.

It might be reasoned that the difficulty with the preceding example could be removed if income were excised as an explanatory variable from (2.2). The expression would not then have any pecuniary variables in it and would therefore seem amenable to the customary epidemiological ministrations. These customary ministrations would, however, continue to be incorrect, for the individual is able to influence the quantity of cigarettes, Z, that he smokes during the current period. If air pollution exposures change, the individual is likely to change the quantity of cigarettes that he smokes. Thus, even after excising the income variable from (2.2), possibilities for biasing the air pollution coefficient remain. To see this, write:

$$\P_i = \alpha_1 + \alpha_2 E + \alpha_3 X + \alpha_4 Z + \varepsilon_1. \tag{2.5}$$

$$Z = \beta_1 + \beta_2 \P_i + \beta_3 P_Z + \beta_4 P_k + \beta_5 Y + \varepsilon_2. \tag{2.6}$$

The variables in expression (2.5) are defined as in (2.2). Expression (2.6) states that the quantity of cigarettes the individual currently smokes is a linear function respectively of whether or not he has chronic bronchitis, the price of cigarettes, the prices of goods that are complements and/or substitutes for cigarettes, and his income.

Upon solving (2.5) and (2.6) for $\P_i$, the coefficient attached to air pollution, E, proves to be $\alpha_2 \left[1 - \alpha_4 \beta_2\right]$, which represents a mix of effects due to air pollution, cigarette smoking, and the effect of bronchitis on cigarette smoking. Again, the product of the coefficients for the latter two effects would have to approach zero for the response of bronchitis to air pollution alone to be obtained. In addition, the sign of the E-coefficient would again be ambiguous since $\beta_2 \gtrless 0$. If $\beta_2 > 0$, the effect of air pollution would be [illegible]erestimated, and if $\beta_2 < 0$, the effect would be underestimated.

To attempt to account for the additional factors thought to influence a morbidity or mortality rate by simply stringing out variables in a single expression must clearly often be incorrect. During the period in which the health effect is supposed to occur, humans acting in their individual capacities can choose to influence the magnitudes assumed by certain of these variables. Each variable susceptible to this influence must be explained by an expression of its own. Economic analysis is necessary to impart an interpretable form to these expressions. Physical and biological constructs will therefore often be insufficient tools with which to provide epidemiological explanations of disease incidences.

The previous two examples are about problems of joint determination which involve economic variables. Nevertheless, the problem of joint determination does not require the presence of economic variables. For example, epidemiological studies frequently group disease incidences by individual city and employ measures of central tendency of incidence and other variables as single units of observation. Thus one might try to explain the frequency of deaths from cancer in a sample of U.S. cities by relating it to the dietary habits, air pollution exposures, and median age of the population in each city. Among the dietary variables, one might include saturated fats and cholesterol, dietary components frequently said to be positively related to cardiovascular disease. Inclusion of these two variables in an expression intended to estimate the factors that contribute to cancer incidence would probably result in negative signs being attached to their coefficients, implying that saturated fats and cholesterol prevent cancer. This may, in fact, be true, but only indirectly. Specifically, median age in each city will tend to vary inversely with the incidence of cardiovascular mortality; in other words, earlier death reduces median age. Thus, since cancer incidence is positively influenced by median age, one might expect cancer to exhibit negative associations with saturated fats and cholesterol even if they have no direct causal relationship with cancer incidence. The apparent effects of these two dietary variables upon cancer incidence would actually represent a confounding of: (1) the effect of the two variables upon cardiovascular disease; (2) the relation between cardiovascular disease and median age; and (3) finally, but only via (1) and (2), the effect of the two variables upon cancer incidence. In short, at least one other expression explaining median age is required.

2.4 The Costs of Pollution-Induced Disease

The preceding sections have discussed the circumstances under which microeconomics and its methods of empirical application can contribute to the epidemiology of pollution. It was observed that in trying to establish dose-response functions for particular pollutants, it is necessary to be extremely sensitive to the presence of jointly determined variables. Failure to account properly for these variables in the structure to be estimated can result in badly distorted depictions of the effect of a health input such as pollution upon the output, the state-of-health or the incidence of a particular disease. One could, of course, consider all variables to be endogenously determined in some ultimate sense. The key to stopping short of including the entire universe in the structure to be estimated is the formation of intelligent judgments about those variables important to the question of interest over which the individual or system (e.g., urban areas) can immediately exercise no more than trivial control. The number of expressions must equal the number of variables it is posited that the individual or system can control if a determinant solution is to emerge. Most importantly for our purposes, since many of the jointly determined variables in a dose-response structure will be economic requiring the application of microeconomic analysis in order to specify how they are to be introduced to the structure, the actual design of epidemiological studies must often include microeconomic considerations.

The potential application of microeconomic analysis to epidemiological concerns extends beyond the estimation of dose-response functions. The analysis can be used to establish pecuniary values for pollution-induced health effects. These values, which are consistent with the axiomatic structure of benefit-cost analysis, can contribute to evaluations of the economic efficacy of existing and proposed pollution control programs. Attempts to establish these values can adopt two polar views of the individual's degree of comprehension of the relation between pollution and his state-of-health.

The first of these views presumes that the individual fails to comprehend any connection between pollution and his health state, even though pollution does influence this state. To obtain the total loss due to a pollution-induced health effect, this view justifies the estimation of a dose-response function and the multiplication of the loss in health attributed to pollution by a pecuniary value for the health loss. The information and criteria used to set the pecuniary value, and thus the total pecuniary loss, come from outside the system being studied. The basic presumption is that the individual is unaware of the health effects of pollution and therefore does not make any voluntary adjustments in response to its presence.

In addition to being a relatively easy and therefore desirable way to establish pecuniary values for health losses, this first view has the further advantage of reducing the force of the joint determination problem. It thus removes problems similar to the cigarette example of the previous section, where, in response to the presence of increased air pollution, the individual chose to reduce his cigarette consumption. However, the view would affect neither the income nor the dietary examples, for the ill-health caused by pollution can affect the individual's earnings capacity and his dietary habits. These earnings and habits would therefore change as pollution changes, even though the individual is utterly unaware of the cause and, consequently, fails to make any behavioral adjustments in response to pollution.

The polar opposite of the above view is that the individual is fully cognizant of the health effects of pollution and continually adjusts his voluntary behavior accordingly; that is, given the opportunities he has available and the relative prices he faces, he alters his behavior so as to minimize the value of the pollution-induced health losses he suffers. These voluntary adjustments will involve shifts in his time and budget allocations such as reductions in the time and intensity of outdoor activities, pursuit of a less toxic diet, and more visits to the family physician. A view of the individual that presumes he is unaware of the health effects of pollution does not account for these adjustments. In effect, it assumes that, whatever the variations in pollution, the individual's time and budget allocations have always accorded with the allocations occurring at the time of observation. Since, according to the second view of the individual's response to pollution variations, these observed allocations are the result of attempts to mitigate the health effects of pollution, the first view of the individual results in underestimates of pollution health effects. Furthermore, if individuals do reallocate their time and their budgets in response to pollution variations, then measures can be obtained of the income the individual would have to receive or would be willing to pay to leave himself as well off as he was before a change in pollution. These measures correspond to the ideal measures of economic loss established in the microeconomic theory of consumer behavior.

Chapter III

SOURCES OF ERROR

3.1 Problems in Statistical Analysis

The previous chapter introduced the problem of joint determination of many variables - especially those which involve choice by individuals - in epidemiological relationships. This problem, if not explicitly accounted for, can introduce simultaneous equation bias. Estimated effects will not approximate actual (population) values. In other words, even for large samples (those approaching infinity) estimated coefficients are no longer consistent; they do not approach their true population values. A number of techniques are available for providing consistent estimates in simultaneous equations. One of these is described in 4.3 below and the technique is applied both in the Sixty-City experiment, Section 4.5, and in the Michigan Survey experiment, Section 5.6. This chapter thus addresses a number of remaining statistical obstacles to obtaining unbiased estimates and significances of the effects of air quality on human health.

3.2 Heteroskedasticity

Any empirical exercise involves error. To act otherwise is to fool one's self, if not the reader. The error can be due to an inability to capture all the a priori factors that influence the phenomenon of interest, it can be caused by failures in measuring the magnitudes of the variables one has a priori grounds for introducing, or it may be a consequence of a misunderstanding of the structure of the phenomenon. In addition to altering the estimated values of coefficients and/or confidence intervals, errors are registered in the constant terms and the residuals of estimated expressions. The so-called statistical "classical linear model," which is employed to establish all the relations of this volume, presumes that the mean of the error variance (a measure of the dispersion of the observations of the magnitudes of a variable around its average magnitude) is equal to zero. This implies that the errors are constant for observations on all basic units of analysis.

In our mortality study, if the unexplained portion of the incidence of cancer-induced death tends to increase with the size of city, then the error will not be constant from one observation to another. Similarly, in our morbidity study, if the unexplained portion of the duration of chronic illness increases with the value of some variable, then we have again violated a basic premise of the classical linear model. Thus, for example, one might reasonably expect that in locations where air pollution is low and that the variation around this average level would not be very great. Low concentrations of air pollution are unlikely to generate severe chronic illnesses of long duration. However, where air pollution concentrations are high, both the average level of air pollution-induced chronic illness and the variations around this average are likely to be substantial. In low pollution locations, even those with a biological propensity to be harmed from pollution do not suffer any ill effects. However, those with this propensity might be struck down if they are moved to a high pollution location, whereas those who have great resistance will suffer little, if at all. The variation in the duration of chronic illness is therefore much higher where pollution is suffocating because the magnitude of the greatest suffering has greatly expanded, while the magnitude of the least suffering continues to be zero.

Nonconstantcy of the variances of the errors (residuals) in an estimated expression is termed "heteroskedasticity," a term the linguistic roots of which we don't know. Because it means that variation in the errors of an expression varies systematically over observations, it implies that the confidence intervals for estimated coefficients will also vary systematically. The result is that the same basis will not be used to calculate the confidence intervals among observations. Thus, although the estimated coefficients are not affected, the standard errors of these coefficients will be biased. As a consequence, the customary tests of significance have no meaning. Nevertheless, if one knows the direction of the bias, one can sometimes ascertain whether these customary tests of significance accord excessive or too little precision to the estimated coefficients. For example, Kmenta (1971, p. 256) provides a formula that under limited circumstances, permits the calculation of this magnitude and the sign of this bias in standard errors. He also outlines ways in which the raw data can be corrected to negate heteroskedasticity.

3.3 Multicollinearity

Multicollinearity occurs when two or more explanatory variables are so highly correlated among themselves that it becomes difficult to separate or determine the independent effect of each variable. In the extreme case where two variables are perfectly collinear, they are effectively identical. However, if *a priori* information exists on the effect of the collinear variables, then that information can be used. For example, if in attempting to explain stomach cancer mortality rates using cross-sectional data, two explanatory variables, sulfur oxides in air and per capita consumption of Polish sausage, are perfectly collinear, one might employ data from animal experiments or epidemiological studies on select human populations (e.g., Polish populations and industrial workers exposed to SO_2 in high concentrations) to determine the relative incidence of stomach cancer from each factor. By including only one of the variables in the regression, the total effect of both explanatory variables will be captured by the estimated coefficient on that one variable. Thus, if consumption of Polish sausage and sulfur oxide exposures are perfectly collinear and only consumption of Polish sausage is included in the estimated equation, the estimated coefficient on consumption of Polish sausage will capture the effect of both variables. How that effect is to be allocated between the two variables depends on the availability of external information. For example, if animal experiments do not show a link between sulfur oxide exposures and stomach cancer, but do show a link between consumption of cured meats (including Polish sausage) and cancer, one might allocate the entire coefficient to consumption of Polish sausage. Of course, if this were the case, and the investigator did not know that consumption of Polish sausage and sulfur oxide exposures was perfectly collinear and no dietary data was available for inclusion, then a false link between sulfur oxides and stomach cancer might be shown using the cross-sectional data alone.

The same arguments apply to cases of near perfect multicollinearity wherein explanatory variables are highly, as opposed to perfectly, correlated. This is, of course, the most likely case. However, the outcome of including two or more collinear explanatory variables is an increase in the standard error of the estimated coefficients for the collinear variables. The standard error is, of course, a measure of the accuracy with which a coefficient is estimated -- large standard errors imply that the actual coefficient could be much larger or smaller than the estimated coefficient. Thus, when collinear variables are included, the inability to separate influences is reflected in the measure of uncertainty over the magnitude of the estimated coefficients on those variables.

The approach taken here to deal with multicollinearity -- and the 60-city experiment described below has a severe problem among the dietary variables -- is to a priori exclude variables which are highly collinear with respect to a representative included variable. An alternative approach to multicollinearity is the use of a technique known as ridge regression [see Schwing, et. al. (1974)] which, however, makes interpretation of the resultant estimated coefficients unclear.

While multicollinearity within an available data set makes estimation and interpretation more difficult, at least the problem can sometimes be recognized and false conclusions thereby avoided. However, where unknown collinearity occurs, for example when an included explanatory variable is highly collinear with a variable which is not available to the investigator, the false conclusion can be reached that the included variable is solely responsible for the estimated effect. The investigator may not recognize that the estimated effect includes the effect of one or several other excluded but collinear variables. We discuss this possibility below.

3.4 Causality and Hypothesis Testing

Aside from the problem of multicollinearity, the traditional problems of causality underlying epidemiological studies still apply. For example, if heart attacks are actually related to cigarette consumption, but smoking is correlated with coffee consumption for behavioral reasons, a spurious positive correlation might be shown between heart attacks and coffee consumption, especially if cigarette consumption is excluded from an estimated statistical relationship. In other words, correlation does not prove causation, and statistical hypothesis testing can never confirm, but only reject, a maintained hypothesis. Turning to another example, if most nitrite (used to cure meats) ingestion is through consumption of pork products (70 % of pork is cured), one might suspect, given the hypothesis of in vivo nitrosamine (a carcinogen) formation from nitrite, that cancer mortality and pork consumption would be correlated. If such a correlation can be shown (as it has been; see Kneese and Schulze (1977) and NAS 1978) then the only valid conclusion is that we do not reject the hypothesis that pork consumption (and perhaps, in turn, nitrite ingestion) is related to human cancer. If, alternatively, one accepts the maintained hypothesis on a priori grounds, and no bias exists in the estimation procedure, regression analysis can give a best linear estimate of the actual relationship in the sample population between, for example, cancer mortality and a dietary factor such as nitrite ingestion. However, regression analysis cannot prove causality; causality must be assumed in this procedure. This is why it is so important to have hypotheses concerning causality before a regression equation is specified.

A set of hypotheses concerning human health, including the effect of air pollution, forms a model of human health. The concept of a complete model of human health as the basis for hypothesis testing is an important one for several reasons. First, a modeling framework immediately suggests that behavioral elements such as voluntary medical care may be important and as pointed out above, a simultaneous equation structure may be necessary to test hypotheses properly. Second, the modeling framework focuses attention on a complete specification of the determinants of human health. A "better" model will exclude fewer relevant variables and be both a more accurate predictor of human health and more accurately identify the effect of each explanatory variable. The modeling approach then helps avoid the problem of unknown collinearity by focusing on a specification which provides information about the effects of all relevant variables.

An alternative viewpoint has been expressed by Lave and Seskin (1977). Their argument rests on the assumption that excluded variables (medical care, diet, and smoking are excluded from their study of air pollution and human health) will not bias estimated effects of included variables if the excluded variables are orthogonal (perfectly non-collinear) with respect to the included variables. Thus, if one assumes orthogonality with respect to excluded variables, following Lave and Seskin (1977), one can justify estimation of incompletely specified equations. We take a different approach principally because we reject orthogonality as a reasonable assumption. If, as ecologists are fond of saying, "everything depends on everything else," then simultaneity and collinearity are likely to be pervasive in the "real world." In fact, we argue below based on our own epidemiological and economic data that this is just the case.

Finally, to test specific hypotheses, we will use the standard significance test; we will test the hypothesis that each explanatory variable has no effect (has a coefficient of zero) by using the appropriate t-statistic which, in this case, is approximately equal to the estimated coefficient divided by its own standard error. For example, for large samples, if for a specific coefficient $t \geq 2.0$ (if the coefficient is greater than or equal to twice its own standard error), then, where the hypothesis tested includes an assumed sign for the coefficient, a 97.5% level of significance is achieved. This implies that, in random sampling of a population, one would draw a sample which accidentally confirmed the hypothesis (effect non-zero) only 2.5% of the time.

It is important to note, that as the significance level is implicitly lowered from t = 2.0 toward t = 1.0, even in large samples, spurious relationships begin not to be rejected. Practical experience and econometric tradition suggest that a 95% to 97.5% significance level is appropriate. The desired confidence level should be chosen a priori to avoid the temptation to "prove" desired relationships by ex post lowering of the level of significance for rejecting or failing to reject hypotheses. Similarly, statements that an explanatory variable is "nearly significant" should be interpreted with great caution. Where costly environmental programs are to be justified by epidemiological analysis, rigorous tests of significance should be employed.

3.5 Aggregation

In one or another of his many books, Herbert Simon has used the term "bounded rationality" with reference to limited human abilities to arrange, comprehend, and manipulate large volumes of information. More succinctly, Simon is referring to the need to simplify in order to understand. Even the pure theorist, in both his analysis and exposition, must partition the universe into two parts: that with which he will and won't deal. Moreover, he must employ a limited and often quite small number of concepts to deal with the part he has chosen. He who would measure as well as theorize must simplify beyond this, for he must be economic with his data manipulations. Both isomorphism with his theoretical variables and his less than fully robust empirical tools require this. Simplification is synonymous with throwing away information, but that which is thrown away is often beyond our powers of use. As Stigler (1967) has remarked, ". . . information costs are the costs of transportation from ignorance to omniscience and seldom can a trader afford to take the entire trip."

In the material to follow, we have played the role of the aforementioned trader in two ways. First, in the mortality study, we have employed grouped data for estimation; that is, we have employed a single measure of central tendency (usually the arithmetic mean) of the distribution of some attribute across a group of people or locations (a city) as the sole representation of the group's diversity. We have melted entire cities into one pot. Here we wish to discuss the issues this poses for estimation.

A second aggregation thing we have done is to embrace the notorious representative individual when discussing the pecuniary benefits or costs of a given health effect. Too fond an embrace of this representative can lead to gross errors if his responses are incautiously applied to flesh and blood individuals. We wish to explain why. Initially, however, we will discuss the estimation issue.

In the mortality study, the unit of analysis is a city or some larger jurisdictional unit and the values attached to a particular variable represent the per capita magnitudes of the variable in the cities. To form these per capita magnitudes, someone had to collect observations on the values of the variables for the distinct individuals in each city. By using the per capita rather than variation of the individual observations within each city and thereby reducing the efficiency of our estimators. Simultaneously, we are lessening the degrees of freedom and, thus, the variety of statistical tests we can potentially employ. Our real gain from this is a drastic shrinking of the size of the data base we must manage. A vacuous gain also exists.

By using the per capita magnitudes for the values of our variables, we have not changed the underlying sample of individual observations, but have reduced the variability of the sample we are using for estimation. We have stripped the outlying individual observations of influence. The result is that the per capita magnitudes will be less dispersed around any expression we estimate, allowing us to appear to explain a larger proportion of the variation in the sample; that is, the magnitude of the coefficient of determination (R^2) is enhanced. This enhancement, however, is misleading since it is entirely due to our prior exercise of collapsing all the variations of individuals' observations in a city to a single scaler measure. Similarly, nonvacuously, and therefore much more importantly, by reducing the variation in the sample, we are reducing the standard errors of each estimated (and still unbiased) explanatory variable coefficient. As a consequence, we may be overstating the level of significance to be attached to these coefficients.

Yet another nonvacuous and altogether serious way exists for the estimates obtained from per capita data to be seriously misleading. The measurement unit one is using for any particular variable may differ from city to city. Thus, for example, one might be measuring cigarette consumption per capita in the equivalent of packs in one city and pounds in another. Consider the following simple algebraic argument.

Assume that a disaggregated dose-response expression for respiratory disease is to be estimated. Let this expression be given by:

$$c_{ij} = a_i + b_i p_{ij} + \varepsilon_{ij}, \quad \begin{array}{l} i = 1, \ldots, n \\ j = 1, \ldots, r \\ n \geq r \end{array} \tag{3.1}$$

where i refers to a particular pollutant, j to a particular individual, a and b are coefficients to be estimated, and ε is an error term having the customary properties. Per capita responses and doses are clearly:

$$C_j = \frac{\Sigma_i \, c_{ij}}{n} \tag{3.2}$$

$$P_j = \frac{\Sigma_i \, p_{ij}}{n} \tag{3.3}$$

With aggregation, the intercept and error terms are:

$$a = \frac{\Sigma_i \, a_i}{n} \tag{3.4}$$

$$E = \frac{\Sigma_i \, \varepsilon_{ij}}{n} \tag{3.5}$$

The aggregate relation is therefore:

$$Cj = a + bP_j + E_j, \tag{3.6}$$

where b, the coefficient of P_j, is apparently

$$b = \frac{\Sigma_i \, b_i}{n} \tag{3.7}$$

In other words, the per capita response depends on the exposures suffered by the n individuals. This perhaps seems reasonable, since (3.6) continues to be linear and includes an error term the expected value of which is zero for all c_{ij}.

Disregarding a and E_j, note, however, that both b and P_j are aggregated. Thus:

$$bP_j = \frac{(\Sigma_i \, b_i)}{n} \frac{(Pi)}{n} = \frac{\Sigma_i \, b_i \, p_{ij}}{n} , \tag{3.8}$$

and therefore

$$b = \frac{\Sigma_i b_i p_{ij}}{nP_j} . \tag{3.9}$$

Nothing goes awry if the dose-response functions are identical among sufferers. However, if they do differ, it is apparent from (3.8) that the value of the pollution exposure parameter, b, will be a weighted mean of the same parameter for the individual suffers. In particular, those sufferers having high responses will have a disproportionately strong influence upon a group's (e.g., a city) contribution to the value of the exposure parameter in (3.8). Similarly, those groups having low responses will have a disproportionately weak influence. The conclusion is the rather dismaying one that the measure of responses, employing some group or aggregation of individuals as the fundamental unit of observation, can differ from one group to another. There could conceivably be as many unique measures employed as there are groups.

The preceding remarks refer to the prior aggregation of individual observations and the subsequent use of the aggregates for estimation purposes. Suppose we employ individual observations for estimation purposes, establish responses for the representative individual among these observations, and then use the presumedly representative responses of this representative individual to obtain an aggregate measure of total response; that is, we aggregate after rather than before estimation. The study of the morbidity effects of air pollution that follows readily lends itself to this treatment. Because it does so, we feel it worthwhile to caution the reader about the dangers this form of aggregation poses. We state the discussion in terms of demand functions although dose-response functions would serve equally well. Only because it is perhaps the most widely cited study to aggregate individual observations of air pollution damages, we employ Waddell (1974) as a basis for discussion.

Waddell (1974) first reviewed a collection of studies that had estimated marginal purchase price functions with respect to sulfur oxides and/or suspended particulates for eight different cities. Interpreting the values of the air quality parameters in these several studies as measures for the average household in each study of equilibrium marginal willingness to pay at given air quality states and with given demand functions for air quality, he selected a value within the range of these estimated values. By selecting this value within the range of values, he assumed that what was interpreted as the equilibrium marginal willingness to pay was the same for all household in all cities.

Then, using the further assumption that this assumed equilibrium marginal willingness to pay was in fact the actual marginal willingness to pay for all air quality states, he multiplied the constant marginal willingness to pay by the number of households and the number of air quality states to obtain an estimate of aggregate national air pollution damages. That is, if b is the marginal willingness to pay and Q is an air quality state, Waddell (1974) calculated aggregate national air pollution damages, D, as

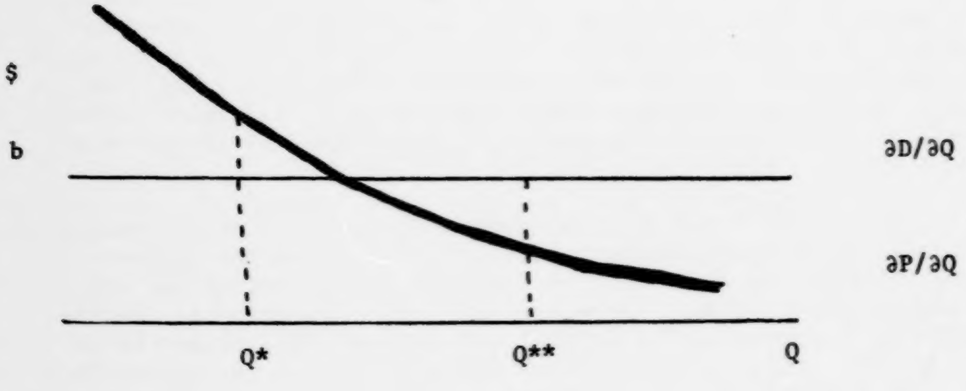
$$D = b\sum_{i=1}^{n}\Delta Q_i \qquad (3.10)$$

where the i's index households.

In effect, Waddell (1974) assumed that the decision problem of each and every household in each urban area of the country could be represented as depicted in Figure 3.1. In Figure 3.1, $\partial P/\partial Q$ is the marginal purchase price function and $\partial D/\partial Q$ is the function representing marginal willingness to pay for improvements in air quality. Since $\frac{\partial D}{\partial Q} = b$ is invariant with respect to changes in air quality, calculation of that willingness to pay for the household of Figure 3.1 involves only the multiplication of b by whatever change in air quality is postulated. Thus, the value to the depicted household of an improvement in air quality ($Q^{**} - Q^*$) is simply $b(Q^{**} - Q^*)$. Given then that b is the same and invariant for all households, the sole distinction one need make among households in order to calculate aggregate national damages is to account for the location of each household on the Q axis.

Figure 3.1

Marginal Purchase Price and
Marginal Willingness-to-Pay



Among the more significant i.e., stronger assumptions in the aforementioned calculation procedure are the following. First, it is assumed in the procedure that the b's are invariant across households. By dropping this assumption, the immediately preceding expression becomes:

$$D = b_i \Sigma \Delta Q_i. \tag{3.11}$$

This would mean that differences in willingness to pay for improvements in air quality due to differences among households in such personal attributes as income, age, and degree of risk aversion to health effects would now be taken into account. Aggregation would then not entirely destroy knowledge about relative sufferer valuations of alternatives.

A further weakening of assumptions would occur if the marginal willingness to pay function is permitted to be nonconstant and even nonlinear. In this case, the above expression for D would be:

$$D = \Sigma_i \int_{Q^*}^{Q^{**}} b_i(Q_i)dQ_i \tag{3.12}$$

Clearly, this would be a very complex expression with which to calculate aggregate national air pollution damages. Not only are the marginal valuations of given air quality states permitted to vary among households but the responses of different households to similar variations in air quality are also permitted to differ. The sensitivity of the aggregation procedure to differences in the economic and air pollution circumstances of households would be greatly enhanced. Freeman (1974, pp. 81-82) lists several frameworks for constructing algorithms that might approximate this last expression for D.

The above discussion has been devoted to a single aggregation over individual households. It has been implicitly presumed that only a single class of air pollutants is relevant. Typically, however, estimates of national air pollution control benefits involve aggregation over multiple classes of pollutants as well as over households. On occasion, aggregation may, in addition, take place over time. Scaler estimates of the national benefits of air pollution control may thus involve two or three distinct types of aggregation, each of which embodies unique assumptions about the similarities among the units undergoing aggregation.

An additional decision problem, over and above the problem involving the manner in which the units in each type of aggregation are to be treated as similar, is thus introduced: one must choose thich type of aggregation is to be performed first in arriving at a scaler representing air pollution control benefits for households, for pollutants, and for time intervals. Moreover, in deciding how to perform the first aggregation, one must take into account how the aggregation for the second and third steps will be carried out. The order in which the aggregation is performed can make a difference in the estimate one obtains of aggregate national benefits.

Chapter IV

THE SIXTY-CITY EXPERIMENT

4.1 Objectives of the Experiment

Identification of substances in the environment which effect human health and accurate quantifications of their effects, is extremely difficult. Often there are multiple substances involved, there may be long latency periods before effects are seen, and the amount and time of exposureis often unknown. There are three general approaches to identifying such substances and quantifying their impact -- all more-or-less imperfect. In the first, laboratory animals are exposed to large doses of the suspect substance and, if effects appear, an effort is made to extrapolate them to the human population. The correct manner in which to execute the second step is not well extablished. The second approach is to develop aggregate cross-sectional data, usually for cities or standard metropolitan areas, on a number of variables which might be associated with mortality rates or illness rates and then to use regression analysis in order to discover statistically significant associations. A third approach is to gather very detailed data on individuals and to again use statistical analysis to determine the effect of various factors including environmental exposures on individualized measures of health status.

The purpose of the research reported in this chapter is to explore both the possibilities and limitations of the second approach mentioned above --aggregate epidemiology -- in the estimation of human dose-response functions which include exposure to air pollution. The principal advantage of the use of aggregated data on cities or metropolitan areas is quite simply the widespread availability and low cost of such data as opposed to data generated from animal experiments or collected on individual human beings through specialized surveys. However, the use of aggregated data creates a number of special problems.

First, one ideally wishes to estimate a dose-response relationship or function as shown in Figure 4.1. Based on a priori considerations one would suppose that for human populations, risk of death for an individual would be a function of medical care, age of the individual, the genetic endowment of the individual, the behavior of the individual--does he or she exercise, smoke, etc.--the diet of the individual, and exposures to possibly harmful substances or circumstances. However, aggregate epidemiology provides no data on individual risks or characteristics but only data for population characteristics as a whole. Thus, aggregate mortality rates in, for example a city are used as a proxy for risk of death in the estimation of an individual dose-response function where it is implicitly assumed that individuals can be represented by the average individual in each city. Thus, in using the data set developed below for sixty U.S. cities to estimate a dose-response function of the form shown in Figure 4.1, it is implicitly assumed that each city represents one average individual. However, the list of assumptions required to allow such aggregation (all relationships must be perfectly linear, etc.) are not likely to be met in practice. Thus, one must recognize that estimated results are biased to an unknown extent by the very use of aggregated data.

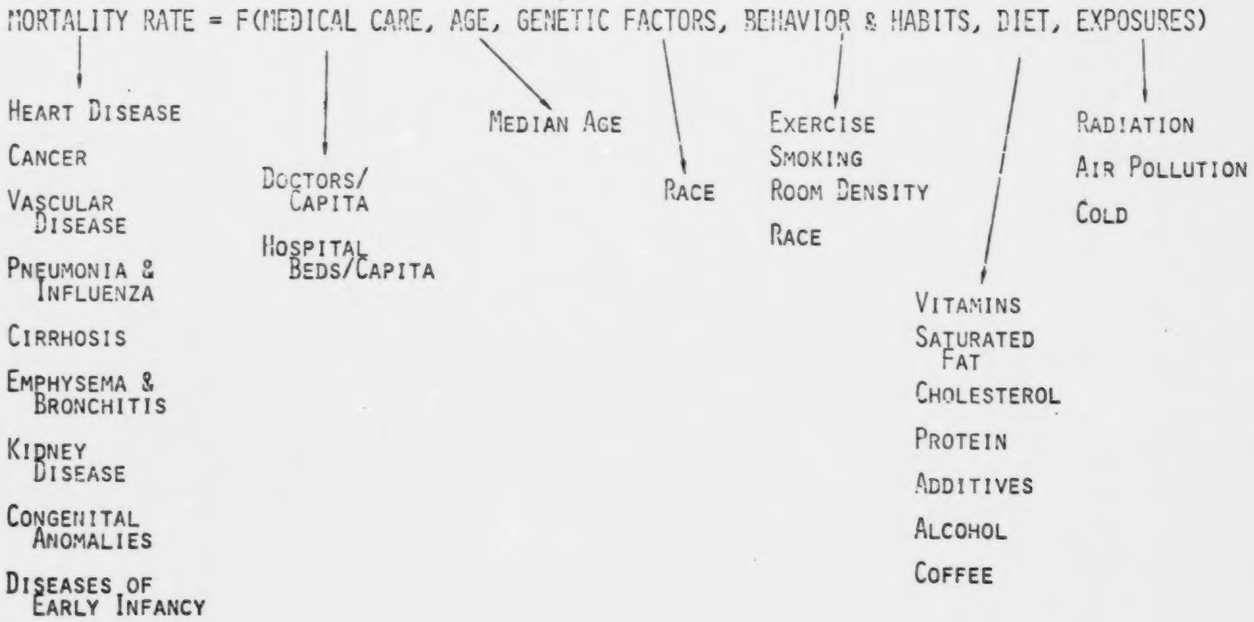
A second problem arises from the fact that aggregate epidemiology must rely on secondary data. Since the investigator must depend on data already collected, he cannot add a question to a survey nor can he vary the design of an animal experiment to test the importance of a new variable. In the past this problem has led to the exclusion of data on important variables such as smoking, diet and exercise from aggregate epidemiological studies [see, for example, Lave and Seskin (1977) and Schwing, et. al. (1974)] We have been able to gather some data -- not necessarily good data -- on both smoking and diet and as we show below, these are important omissions from previous studies. The current study still excludes any measure of exercise.

Finally, aggregate epidemiological studies are likely to suffer from a number of simultaneous equation biases. One of the most obvious concerns the effect of medical care. The existing epidemiological literature has failed to show any significant effect of medical care on human mortality rates. This counterintuitive result is easily explained. For example, in our sixty city sample, no effect of per capita doctors in each city can be shown on aggregate mortality rates for each city when simple regression techniques such as ordinary least squares are used. The explanation is that, although doctors most likely do reduce mortality rates (as shown below), people in cities with higher mortality rates have in turn more illness per capita and seek out more medical care, increasing the observed number of doctors in such communities. In other words, higher mortality rates create a greater demand for doctors. Thus, we have two offsetting effects--doctors reduce mortality, while mortality increases the demand for doctors--and simple regression analyses cannot untangle them. Several statistical techniques are available for coping with simultaneous equation problems.We use a very simple approach, two-stage least squares, a technique described in a little detail below.

A second simultaneous equation problem may arise because of multiple causes of death. Cities with high coronary death rates may likely have lower cancer death rates because people die of heart attacks before they have a chance to die of cancer. In this situation, factors which, for example, show up positively correlated with coronary disease may show up with a spurious negative correlation with cancer rates. This simultaneous equation problem is likely to w[illegible] "through" the age variable in that median age is determined in [illegible] mortality rates of individual diseases, while, in turn, age is used [illegible]lain mortality rates. We therefore have also employed two-stage least squares on the age variable, but with no impact on the estimated equations so these results are not reported here.

Figure 4.1

Hypothetical Human Dose-Response Function



An alternative approach to the problem which we do not employ, is use of age-specific mortality rates.

A third possible source of simultaneous equation bias occurs because people make voluntary choices over location. Migration in and out of our sixty city sample is effectively disregarded. People may, for example, contract an air pollution related disease and, on a physician's advice, move from a highly polluted area to an unpolluted area, only then to die. A false negative association between air pollution and pollution-related mortality might then be shown. Although in the past we have included a net migration variable [see Kneese and Schulze (1977)] which was statistically significant, we have excluded such a variable in this analysis because it defies interpretation in a dose-response function context.

Table 4.1 summarizes the objectives and limitations of the current study and to some extent those of aggregate epidemiology in general. We now turn to development of methodology for estimating the value of reducing health risks and for the effect of medical care on human health. This latter section focuses on the role of exonomics in aggregate epidemiology.

4.2 Value of Life Vs. Value of Safety

The direct costing approach employed by economists for evaluating the mortality costs of diseases which result from environmental exposures is straightforward but difficult to quantify fully [see, for example, Kneese and Schulze, 1977]. First, the population at risk must be known. Second, the increased risk of mortality associated with environmental exposures must be quantified either through epidemiology or through extrapolation from animal experiments. Third, the amount of money or the value that individuals place on safety (avoiding risk of death) must be known. Multiplying these three values together then gives an approximation of the incremental benefits of reducing such exposures. This cost or benefit is not in any way related to a "value of life" which is most likely unmeasureable, but rather focuses on a concept of the value of safety (alternatively "cost of risk") to individuals where risks are statistically small.

Table 4.1

Objectives and limitations

Purposes of Study Are:

(1) To explore methodology for isolating an aggregate human dose-response function.

(2) to add medical inputs.

(3) to add diet.

(4) to add smoking.

(5) to account for simultaneous equation bias where possible including:

  (a) demand for doctors.

  (b) multiple possible causes of death.

The Study Fails to Account for:

(1) simultaneous equation problems caused by migration.

(2) exercise.

(3) radiation.

(4) Biases introduced by estimating an aggregate as opposed to individual dose-response function.

Economists in the past have attempted to value human life as the sum of the present value of future earnings over an individual's lifetime [see Lave and Seskin, 1970 and 1977]. This approach, however, is no longer viewed as acceptable. In the first place, it assumes that the value of life can, in fact, be measured -- a point certainly open to debate. Second, it implies that the lives of children, housewives, retired and other unemployed individuals are worth less than the lives of employed heads of households.

Two measures can be used to value safety or risk to life which are based on the economic concepts of equivalent variation (EV) and compensating variation (CV). An EV measure of the value of life is the amount of money an individual would pay to escape from or prevent certain death; in theory, a rational individual would part with all his available wealth to save his life. CV, in contrast, measures the compensation required to induce an individual to accept voluntarily a situation where the probability of death is increased. As the probability of death approaches unity, the CV measure can be taken as an estimate of the value the individual places on his life. Logically, though, the value of life measured this way must be infinite, because as the probability of death approaches certainty, the probability of enjoying any compensation offered (and thus the value of the compensation) approaches zero. Thus, neither EV (which requires coercion) nor CV (which makes the value of life immeasurable) provides a wholly satisfactory way of extimating the dollar costs of mortality in realy world situations that involve risk. An eleboration of the CV concept, however, can provide a useful measure of the compensation necessary to induce an individual to accept a slight increase in the probability of death.

Mishan (1971) was the first to distinguish between the concept of cost of risk, which is ethically appealing, and earlier efforts to value human life based on lost earnings, which as a methodology, has strange and intuitively objectionable features. The latter measure of the "value" of human life has now been rejected by economists both on theoretical and, to some extent, on ethical grounds. Thaler and Rosen (1975), using wage differentials between jobs varying in the level of job-associated risk of death, were the first to estimate explicitly the value of safety. In other words, workers in high risk jobs receive higher wages and a value of safety can be inputed by examining risk-related wage differentials. Unfortunately, however, their study dealt with a high risk class of individuals. The Thaler and Rosen (1975) estimate suggests that in current dollars a small reduction in risk over a large number of individuals which saves one life is worth about \$340,000. Another study [Blumquist (1977)],which examines seat belt use, suggests that the figure might be \$260,000. This study first estimates how people value their own time and then imputes a value of safety from the amount of time a sample of individuals spent in buckling up seat belts. These results may be biased downward because individuals seem to perceive risks differently when an element of personal control such as driving an automobile is involved. Finally, Smith (1975) in a study similar to Thaler and Rosen (1975) has suggested that, for a more typical population and for job-related risks, the figure may exceed \$1,000,000. Clearly, the cost of risk is not precisely known, and perhaps will never be, since attitudes -- risk preferences -- presumably can change over time, between groups, and can even vary in different situations. But, we at least have a range of values with which to make order-of magnitude estimates of the costs of environmental risks. This range of values does not, however, overlap the value-of-life estimates based on lost earnings. For example, Lave and Seskin (1977) use a value in the thirty to forty thousand dollar range for a life lost. The Thaler and Rosen (1977) value of safety is, for example, about an order of magnitude larger than the Lave and Seskin (1977) lost earnings number.

The theoretical basis of a value of safety or cost of risk concept can be shown briefly as follows: Assume that an individual has a utility function, $U(W)$, where utility is an increasing function of wealth, W. If risk or death is Π, expected utility is $(1-\Pi)U(W)$. If we hold expected utility constant, we have $(1-\Pi)U(W)$ = constant, and the total differential of this equation is:

$$-U(W)d\Pi + (1-\Pi)\,U'(W)dW = 0 \tag{4.1}$$

where the prime denotes differentiation. Holding utility constant then implies that the increase in wealth (or income) necessary to offset an increase in risk is:

$$dW/d\Pi = U/[U'\,(1-\Pi)]. \tag{4.2}$$

This is the compensating variation measure of the cose to an individual attributable to an increased risk of death. Analysis of the last expression can be simplified if we assume a constant elasticity of utility with respect to wealth, η, such that $U(W) = \eta$ and consequently $\eta = \frac{dU}{dW}\frac{W}{U}$. Then (4.2) can be rewritten as:

$$dW/d\Pi = W/[\eta\,(1-\Pi)]. \tag{4.3}$$

The right hand side of (4.3) suggests several interesting points about the value of safety or cost of risk. First, if we assume that the elasticity of utility is less than one, people are risk averse. This in turn implies that since the risk of death is positive ($\Pi>0$) that $(dW/d\Pi)>W$. In other words, if an individual is risk averse, his life, in terms of the risk premium necessary to get him to accept risk, is worth more to him than his wealth. Second, from (4.3), as wealth increases, the risk premium required to accept an increase in risk voluntarily, $dW/d\Pi$ must increase with age, ceteris paribus. Thus, one would expect older people to act in a more risk averse manner than younger individuals (require greater compensation to voluntarily take a risky action), both because of increased income and because of increased initial age-related risk of death.

This model contrasts for a number of reasons with the value of lost earnings approach previously used in economic analysis. First, if lost income itself is the measure, the "value of life" measured through lost earnings obviously cannot exceed wealth [see Conley, 1977]. Second, increased wealth will increase the lost earnings measure as well as the cost of risk measure. However, the cost of risk measure may not increase proportionately if a different utility function is used. Third, the lost earnings measure must decrease with age at some point as individuals get older because the expected remaining earnings must decrease, while the cost of risk, as we argued above, will likely increase. Finally, it is clear from (4.3) that as Π approaches unity, $dW/d\Pi$ approaches infinity. In other words, the compensation required to induce an individual to accept a certainty of death voluntarily is infinite. The lost income measure has no similiar property. Nevertheless, the implication is that small increases in risk may be valued in terms of compensation required to induce individuals to accept such risks voluntarily. Individuals, of course, rationally accept small risks on a daily basis, presumably on the basis of some monetary or psychic return.

Given the analysis above, the current methodology of multiplying value of safety numbers times experimentally or epidemiologically determined enviromental risks can then be justified as follows: assuming a utility function $U(W)$ where W is wealth, if risk of death is Π, the marginal cost of risk, as derived earlier, is $(dW/d\Pi)_{\bar{U}} = U/U'(1-\Pi)$, where $\bar{U}$ is a constant utility level. If risk, Π, is a function of pollution, X, where utility functions are identical for N individuals, one would wish to maximize expected utility,

$$N[1-\Pi(X)]U(W), \tag{4.4}$$

subject to a constraint on total wealth, $\bar{W}$, or income of society

$$\bar{W} - NW - C(X^{o}-X) = 0 \tag{4.5}$$

which is allocated to individual wealth, assumed identical for purposes of exposition, (NW), and costs of controlling environmental pollution from the initial level X^{o}, $[C(X^{o}-X)]$. Noting that $\Pi_X > 0$, and $C' > 0$, the first order conditions are (where λ is the multiplier on (4.5) and L denotes the Lagrangian):

$$\partial L/\partial W = N(1-\Pi)U' - N\lambda = 0$$

$$\partial L/\partial X = -N\Pi_X U - \lambda C' = 0$$

These imply:

$$N\cdot [U/U' (1-\Pi)]\Pi_X = C' \tag{4.6}$$

or that the number of individuals, N, times the marginal cost of risk, $[U/U'(1-\Pi)]$, times the marginal effect of pollution on risk, Π_X, equals the marginal cost of control, C'. Clearly, this model abstracts from many welfare threoretic problems but it does imply that estimation of the left hand side of (4.6) as suggested at the beginning of this section is a legitimate approximation of the incremental benefits of environmental control.

In summary, the direct costing of mortality has the advantage of focusing attention on one positive output of environmental agencies which has clear economic value -- safety. It is important, however, to distinguish between the value of safety to consumers which does have measurable economic value -- environmental agencies may be viewed as selling safety to the public -- as opposed to techniques which claim to measure the value of human life. Benefit-cost arguments for environmental programs should and can rest on demonstrable increases in public safety delivered at costs comparable to what the public is willing to pay for safety, not on claims as to the value of human life. However, the assessment of the risk of mortality associated with environmental exposures such as air pollution--whether based on animal experiments or epidemiological studies -- remains difficult and uncertain and is central to the direct costing methodology. Surprisingly, perhaps, the authors feel there is likelv to be less professional debate as to the economic measure of the dollar value of safety than as to the quantification of environmental health effects. We now turn to the possible role of economic analysis in the epidemiology of air pollution.

4.3 A Methodological Basis: Does Economics Matter?

The question posed above could be rephrased "does rational human behavior matter in the estimation of dose-response functions?" Economists would certainly answer in the affirmative; individuals are likely to respond to illness with numerous ameliorative measures. Clearly, such measures must be accounted its a properly specified dose-response function is to be estimated. What follows is a simplified economic model of human behavior in response to health risks which in turn allows specification of appropriate statistical techniques for estimating a human dose-response relationship.

Let Π denote risk of death for an individual where that risk can be reduced by medical care which we denote D, synonymous with our empirical measure, doctors per capita. Thus, risk can be written as a function, $\Pi(D)$, where $d\Pi/dD = \Pi' < 0$. If the price of medical care is p and income is denoted Y, then utility, U, can be written $U(Y - pD)$, a function of income net of expenditures on medical care, pD. In an uncertain world, economic analysis assumes that an individual will choose to maximize expected utility -- the odds of remaining alive $(1-\Pi)$ times the utility level U -- or

$$[1 - \Pi(D)]\, U(Y - pD), \qquad (4.7)$$

so the first order condition for the quantity of medical care chosen when rearranged is:

$$\frac{U}{(1-\Pi)U'} = \frac{P}{-\Pi'} \qquad (4.8)$$

The term on the left-hand side of (4.8) is easily recongizable from section 4.2 above as the marginal value of safety (or compensation required to voluntarily accept a small increase in risk), while the term on the right is the marginal cost of increased safety through medical care. Thus, this model of human behavior implies that an individual will choose a level of medical care which equates his or her marginal value of safety to the marginal cost of reducing risk through medical care. Of course, an individual's perception of risk and of the ability of medical care to reduce risk of death may be imperfect. However, from (4.8) it is easy to show that individuals who are more risk averse, i.e., those with a large marginal value of safety, will seek more medical care than those who are less risk averse.

An explicit set of functional forms will simplify interpretation. Let us again (as in Section 4.2) assume a constant elasticity of utility with respect to income, η, so $U = (Y - pD)^{\eta}$. Also assume a linear (approximate) dose-response relationship, $\Pi = \Pi_o + \Pi'D$, where $\Pi' < 0$ is now a fixed coefficient and Π_o is a positive constant. Equation (4.8) can then be written as:

$$D = (\frac{\eta}{\Pi'}) - (\frac{1}{\Pi'})\Pi + (\frac{1}{P})Y \qquad (4.9)$$

which is a demand equation for medical care. If we take the supply price of medical care to be fixed $P = P^*$ (infinitely elastic supply of medical care), the individual demand for medical care, doctors per capita for example, is then a linear increasing function of total risk Π, since $(\frac{-1}{\Pi'}) > 0$, and of income Y, since $\frac{1}{P^*} > 0$. Of course, we wish, as our principal objective for policy purposes, to estimate the dose-response function:

$$\Pi = \Pi_o + \Pi'D; \qquad (4.10)$$

in particular, we wish to obtain an unbiased estimate of Π', the effect of medical care on mortality and of the effect of other variables such as air pollution. However, any attempt to directly estimate (4.10) is doomed to failure. This occurs because the equation specified for statistical estimation (equivalent to 4.10 where α_o and α_1 are parameters for estimation)

$$\Pi = \alpha_o + \alpha_1 D + \mu_\Pi \qquad (4.11)$$

has a disturbance term μ_Π which is not independent of D. In other words, μ_Π is correlated with D. This is easy to show if we specify the demand equation for doctors (equivalent to 4.9 above with parameters β_o, β_1 and β_2) as stochastic:

$$D = \beta_o + \beta_1\Pi + \beta_2 Y + \mu_D \qquad (4.12)$$

as well, with a disturbance term μ_D. Now suppose some factor (random) embodied in μ_Π causes Π to rise in (4.11). But if Π rises, by (4.12), D must increase since, from (4.9), $\beta_1 > 0$ and D, through β_1, depends on Π. Thus, D depends on μ_Π through (4.12) and D and μ_Π are correlated. Now, if in estimating (4.11) this correlation is not accounted for, not only will estimates of α_1 and α_0 be biased, but if we had included other factors which affect morality such as diet or pollution in (4.11), coefficients on these variables would be biased as well. It is also true that if simultaneous equation biased is present and not accounted for, it becomes possible that the estimated effect of medical care, α_1, will appear not significantly different from zero or even of the wrong sign (note we assume that $\alpha_1 < 0$; that doctors reduce mortality).

We can break the dependence of D on μ_{Π} by first substituting (4.10) into (4.9), or (4.11) into (4.12), to obtain a reduced form equation for medical care,

$$D + \frac{\eta}{2\Pi'} - \frac{\Pi^o}{2\Pi'} + \frac{1}{2P*}Y \text{ or } D = \gamma_o + \gamma_1 Y + \mu_r \qquad (4.13)$$

where μ_r is the disturbance term in the reduced form. This equation can be legitimately estimated since the income variable is exogenous, determined outside the relevant system of equations, and the endogenous variables D and Π, those determined within the system, do not appear on the right-hand side of (4.13). Now, if we estimate (4.13) and obtain unbiased estimates of the two coefficients $\gamma_o = (\frac{\eta}{2\Pi'} - \frac{\Pi}{2\Pi'})$ and $\gamma_1 = (\frac{1}{2P*})$ we can use these along with data on income, Y, to generate a new variable, estimated medical care, D, where

$$D = \gamma_o + \gamma_1 Y. \qquad (4.14)$$

Note that this new variable, D, generated from data on Y does not depend on μ_{Π} and can be used instead of actual data on D to estimate a dose-response function:

$$\Pi = \alpha_o + \alpha_1 D + \mu_{\Pi}^{\sim} \qquad (4.15)$$

This estimated equation gives a consistent estimate of α_1 or Π'. In fact, if the hypothesis that doctors are both important and effective in reducing mortality rates is correct, α_1 should show up negative and significantly different from zero as estimated in (4.15). Note, however, that if individuals perceive that doctors are effective, they will have a strong incentive to seek medical help when ill, thus making a direct least squares estimate of the effect of medical care as specified in (4.11) impossible.

The procedure we have outlined above, two-stage least squares, has been used successfully in many instances to resolve simultaneous equation problems and has the advantage of requiring minimal additional data. In general, if an unbiased estimate of a structural equation (one equation is a simultaneous system) is desired, one need only use ordinary least squares to estimate each of the endogenous variables as a function of all of the exogenous variables in the model (estimate a set of reduced form equations). Then, using the data on the exogenous variables, an estimated data set for each of the endogenous variables is created. Consistent structural equations can then be obtained by replacing each endogenous variable on the right hand side of a structural equation by its estimated equivalent using ordinary least squares.

4.4 The Sixty-City Data Set: Selection of Variables

In this section, we describe the data set itself and also examine some properties of the data with special emphasis on collinearity and consequent implications on the variety and kinds of hypotheses which can be appropriately tested.

Tables 4.2-4.5 present a listing of the variables available for analysis along with the year of the variable, units, mean, standard deviation (S.D), and sources for the data by number, where the number refers to the listing of sources in Table 4.6. Table 4.2 includes total mortality rate calculated from 1970 data on mortality by city divided by 1970 census population. Disaggregated mortality data by disease category -- heart, vascular, pneumonia and influenze, emphysema and bronchitus, cirrhosis, nephritis and nephosis, congenital anomalies, early infant diseases, and cancer -- were also collected for 1970, and divided by 1970 census population to develop mortality rates; exceptions are the congential anomolies and early infant disease categories which were divided by the number of births in each city for 1970 in order to define an appropriate mortality rate. Mortality data for 1970 were chosen because reliable city population estimates are available for that year as opposed to more recent data requiring use of non-census year city population estimates in place of actual data. The disaggregation of total mortality by disease may not be appropriate. However, only data on city mortality was available, as indicated in Table 4.2.

Table 4.3 describes per capita dietary data by city for the years 1955 and 1965. The procedure used to construct the dietary data sets was somewhat involved. Food consumption estimates were first constructed for each of the 60 cities, using data on a sample of about 3,000 urban households, distributed among eight income brackets, for four regions of the United States, collected by the Department of Agriculture for 1955 and 1965. The results are regionally-specific weighted averages of consumption of various foods by families in each income bracket, multiplied by the fraction of each city's population in each income bracket. Data for specific dietary factors were then generated by multiplying the consumption rates of 49 foods by their respective concentrations of a given substance. A number of additional variables are available from the Department of Agriculture for 1965 as opposed to 1955. These include total protein, total fats, and total carbohydrates. As such, these variables provide a better indication of broad dietary patterns as opposed to the 1955 data set.

Table 4.4 describes our data on socioeconomic, geographic, and smoking variables. The socioeconomic and geographic variables were chosen for their consistency for estimating the aggregate dose-response function hypothesized in previous sections. Both the income and education variables are hypothesized to enter the demand equation for medical care, not the dose-response function. We must therefore employ the two-stage least squares estimation technique outlined above. Doctors per capita was chosen as the best available indication of available medical care, both in terms of

Table 4.2

Mortality Variables

Variable		Year	Units	Mean	S.D.	Sources
Mortality Variables						
MO70	Total Mortality	1970	deaths/1000 pop.	11.283	2.161	(18)(6)
HA70	Heart Disease	1970	"	4.216	1.078	(18)(6)
VA70	Vascular Disease	1970	"	1.566	0.395	(18)(6)
PN70	Pneumonia & Influenza	1970	"	0.375	0.114	(18)(6)
EM70	Emphysema & Bronchitis	1970	"	0.178	0.059	(18)(6)
CI70	Cirrhosis	1970	"	0.238	0.106	(18)(6)
NE70	Nephritis & Nephrosis	1970	"	0.058	0.027	(18)(6)
C/B%	Congenital Anom/Births	1970	%	0.473	0.105	(18)(6)
I/B%	Early Infancy/Births	1970	%	1.294	0.333	(18)(6)
CA70	Cancer Mortality	1970	deaths/1000 pop.	1.958	0.402	(18)(6)

Table 4.3

Dietary Variables

	Variable	Year	Units	Mean	S.D.	Source
	Dietary Variables					
NTRI	Nitrites in Food	1955	g/yr/cap	1.27	0.14	(2)(4)(27)
NTRA	Nitrates in Food	1955	g/yr/cap	69.86	9.05	(2)(4)(27)
SFAT	Saturated Fatty Acids*	1955	g/yr/cap	16220.00	874.65	(2)(3)(4)
PROT	Protein*	1955	g/yr/cap	26557.00	1314.00	(2)(3)(4)
CHOL	Cholesterol*	1955	g/yr/cap	234.81	6.98	(2)(3)(4)
CVIT	Vitamin C**	1955	g/yr/cap	16.96	1.46	(2)(3)(4)
CALO	Calories	1955	kcal/yr/cap	1149.7	56.27	(2)(3)(4)
COFF	Coffee	1955	kg/yr/cap	5.83	.70	(2)(3)(4)
ALCO	Alcohol ($ value)	1955	$/yr/cap	17.30	6.06	(2)(3)(4)
XPRO	Total Protein	1965	g/yr/cap	39845.	706.46	(28)
XFAT	Total Fats	1965	g/yr/cap	57512.	1795.7	(28)
XCAR	Carbohydrates	1965	g/yr/cap	123490.	3623.0	(28)
XASA	Ascorbic Acid	1965	mg/yr/cap	42281.	2364.2	(28)
6NTI	Nitrites in Food	1965	g/yr/cap	1.14	.16	(4)(28)(27)
6NTA	Nitrates in Food	1965	g/yr/cap	52.87	2.47	(4)(28)(27)
6SFT	Saturated Fatty Acids*	1965	g/yr/cap	16315.	976.3	(3)(4)(28)
6PRO	Protein*	1965	g/yr/cap	28128.	1603.4	(3)(4)(28)
6CHL	Cholesterol*	1965	g/yr/cap	219.9	5.80	(3)(4)(28)
6CAL	Calories	1965	kcal/yr/cap	1171.1	27.63	(3)(4)(28)
6CVT	Vitamin C**	1965	g/yr/cap	18.65	1.3	(3)(4)(28)
6COF	Coffee	1965	kg/yr/cap	5.40	.18	(3)(4)(28)
6ALC	Alcohol ($ value)	1965	$/yr/cap	25.97	6.45	

*** Includes only animal products.**

**** Includes only vitamin C content for fruits and vegetables eaten fresh.**

Table 4.4

Social, Economic, Geographic, and Smoking Variables

Variable		Year	Units	Mean	S.D.	Sources
Social, Economic, Geographic						
MDOC	Medical Doctors	1970	M.D.'s/100,000	162.8	54.2	(19)
IN69	Median Income	1969	$/yr/Household	10763.	1060.	(6)
EDUC	Education	1969	%>25 yrs w/H.S. diploma	55.3	7.4	(6)
DENS	Crowding in Homes	1969	%>1.5 persons/ room	0.022	0.013	(8)
COLD	Cold Temperatures	1972	#days temp < 0 0°C.	86.9	47.7	(9)
NONW	Nonwhite Population	1969	Fraction	0.226	0.154	(6)
MAGE	Median Age of Population	1969	Years	28.82	2.74	(6)
Smoking Variables						
C156	Cigarettes	1956	packs/yr/cap†	183.52	26.66	(22)(4)
CI68	Cigarettes	1968	packs/yr/cap††	165.80	23.25	(7)(1)

† Data for states, 1960 census population used.

†† Data for states, 1970 census population used.

Table 4.5

Air Pollution Variables

Variable		Year	Units	Mean	S.D.	Sources
Air Pollution						
SU66	Sulfate	1966	$\mu g/m^3$	10.67	5.44	(20)
AM66	Ammonium	1966	$\mu g/m^3$	1.15	1.42	(20)
NI66	Nitrates	1966	$\mu g/m^3$	1.96	0.68	(20)
PA66	Suspended Particulates	1966	$\mu g/m^3$	114.83	33.97	(20)
NO69	Nitrogen Dioxide	1969	ppm	0.076	0.034	(11)
PA70	Suspended Particulates	1970	$\mu g/m^3$	102.30	30.11	(13)
SO70	Sulfur Dioxide	1970	$\mu g/m^3$	26.92	22.20	(13)
NI70	Nitrate, annual mean	1971-73	$\mu g/m^3$	3.13	0.92	(14)
NI90	Nitrate, 90th %-tile con.	1971-73	$\mu g/m^3$	5.21	1.80	(14)
SU70	Sulfate, annual mean	1971-73	$\mu g/m^3$	10.65	4.01	(14)
SU90	Sulfate, 90th %-tile con.	1971-73	$\mu g/m^3$	17.69	7.62	(14)
LEAD	Lead	1970	$\mu g/m^3$	1.33	0.54	(16)
CO74	Carbon Monoxide	1974	mg/m^3	11.86	3.50	(15)
BETA	Beta Radioactivity	1966	PC_i/m^3	0.261	0.091	(20)

preventative and ameliorative care. Alternative variables such as hospital beds per capita were judged inferior, in that underutilization of hospital facilities is a common problem and adjustments for utilization factors would prove troublesome. Also, if one assumes a less than perfectly elastic supply of medical care, the doctors per capita variable is an appropriate supply side variable in that it reflects patient loads for doctors in a particular city. The possible importance of age and cold temperatures in a dose-response relationship are clear. However, the nonwhite and crowding variables may be more difficult to interpret. The nonwhite variable would ideally control to some extent for genetic variations in the population. Obviously, however, this variable may well proxy for education, poverty, habits, etc. Similarly, crowding would ideally be an indication of possible contagion but may really proxy for poverty, old age, or even race. Thus, the role of these variables should be interpreted with great care.

Cigarette consumption was estimated from cigarette tax revenues for each state in which a sample city was located; the result is thus a statewide average that includes rural populations. Per capita cigarette consumption was estimated using the total state population over 16 years of age both for 1956 and 1968. It should be noted that both our dietary and our smoking variables are quite crude, reflecting problems in utilizing secondary data. However, the possible importance of their effects on human health may justify use of even these measures. We also attempted to develop a measure of total exposure to radiation, collecting data on beta radioactivity in air, terrestrial gamma radiation, and cosmic ray exposures, but have to this point been able to account for only about half of the average individual annual dose associated with medical exposures. As a result, no suitable total exposure variable is available at this time.

Table 4.5 presents the air pollution variables available for testing. All data are annual geometric means for each city unless otherwise specified. Unfortunately, hydrocarbon data was only available for about ten of our sample cities and are excluded for this reason. Finally, as noted above, Table 4.6 presents our data sources.

In summary, data available for testing include: (1) 1970 mortality rates for total mortality and for major disease categories; (2) data on dietary patterns for 1955 and 1965; (3) data on medical doctors and socioeconomic factors for 1970 or a nearby year; (4) data on smoking patterns for 1956 and 1968; and (5) data on air quality for each city for 1970 or a nearby year.

Since only 60 observations are available, we must obviously select a subsample of the available explanatory variables for hypothesis testing. To allow straightforward statistical tests of the significance of estimated coefficients, it is necessary to make the selection of included variables a priori rather than testing each of the variables in various combinations for significance and excluding some on the basis of relative significances. Techniques do exist for testing significance where pre-testing has been employed but the standard t-statistic is no longer applicable.

The first problem in specifying the final data set is a decision on including lagged variables. Given a highly mobile U.S. population, the

Table 4.6

Sources of Data

(1) Advisory Commission on Intergovernmental Relations, State and Local Finances: Significant Features and Suggested Legislation, 1972, Table 120, 1970.

(2) U.S. Department of Agriculture, Household Food Consumption Survey, 1955, Reports No. 2-5.

(3) ________, Composition of Foods: Raw, Processed and Prepared, Watt, Bernice K., and Merrill, Annabell L., Agricultural Handbook No. 8, 1968.

(4) U.S. Department of Commerce, Bureau of the Census, U.S. Census of the Population: 1960.

(5) ________, Cross Migration by County: 1965-1970, Current Population Reports Series P-25, No. 701, issued May 1977.

(6) ________, U.S. Census of Population: 1970, Vol. 1-50.

(7) ________, State Tax Collections: 1968, Series GF 68 No. 1, Tables 7, 9.

(8) ________, 1970 Census of Housing by State.

(9) ________, National Oceanic and Atmospheric Administration, Environmental Data Service, Climatological Data, National Summary: Annual 1972, Vol. 23, No. 13, Asheville, North Carolina.

(10) U.S. Environmental Protection Agency, Natural Radiation Exposure in the United States: 1972, Report No. ORP/SID 72-1, Table A-1, 1974 Reprint.

(11) ________, Air Quality Criteria for Nitrogen Dioxide, No. AP-84, Tables 6-10, January 1971.

(12) ________, Chemical Analysis of Interstate Carrier Water Supply System, PB-257600/7BE April, 1975.

(13) ________, Air Quality Data - 1970 Annual Statistics, EPA-450/2-76-019, October 1976.

(14) ________, Air Quality for Nonmetallic Inorganic Ions, 1971 through 1974: From the National Air Surveillance Networks, EPA-600/4-77-003, January, 1977.

(15) ________, Air Quality Data - 1974 Annual Statistics, EPA 450/2-76-011, August, 1976.

(16) U.S. Environmental Protection Agency, Air Quality Data for Metals 1970 through 1974: From the National Air Surveillance Networks, EPA-600/4-76-041.

Table 4.6

(continued)

(17) Department of Health, Education and Welfare, Public Health Service, National Center for Health Statistics, *Vital Statistics of the United States: 1960*.

(18) ________, *Vital Statistics of the United States: 1970*.

(19) ________, *Health Manpower -- A County and Metropolitan Area Data Book*.

(20) ________, National Air Pollution Control Administration, *Air Quality Data from the National Surveillance Network and Contributing State and Local Networks*, 1966 Edition.

(21) ________, *Vital Statistics of the United States: 1972*.

(22) Tobacco Tax Council, *Cigarette Taxes in the United States, 1956*, Table 15.

(23) *Directory of Medical Specialists*, 1960-71, Marquis - Who's Who, Inc., Chicago, Illinois.

(24) Adams, John A., et. al., eds., *The Natural Radiation Environment II*, Proceedings of the Second International Symposium on the Natural Radiation Environment, Houston, Texas, August 7-11, 1972.

(25) Hickey, John, et. al., *The Development of an Engineering Control Research and Development Plan for Carcinogenic Materials*, U.S. Government Printing Office, Washington, D.C. (1977 in press).

(26) Pazand, Reza, *Environmental Carcinogenesis -- An Economic Analysis of Risk*, PhD. Dissertation, The University of New Mexico, July 1976.

(27) White, Jonathan W., Jr., "Relative Significance of Dietary Sources of Nitrate and Nitrite," *Journal of Agricultural and Food Chemistry*, Vol. 23, No. 5 (1975), Table VI, p. 890.

(28) U.S. Department of Agriculture, *Household Food Consumption Survey*, 1965-66, Reports No. 7-10 and Reports No. 13-16.

Table 4.7

Simple Correlation Matrix for 1965 Diet Variables

	XPRO	XFAT	XCAR	XASA	6NTI	6NTA	6SFT	6PRO	6CHL	6CVT	6COF	6ALC
XPRO	1.00	0.46	-0.01	0.53	-0.64	0.43	0.16	0.29	0.67	0.88	-0.14	0.70
XFAT	0.46	1.00	0.85	-0.17	0.34	0.74	-0.62	-0.41	0.50	0.12	-0.09	-0.28
XCAR	-0.01	0.85	1.00	-0.33	0.66	0.58	-0.66	-0.46	0.33	-0.31	-0.22	-0.71
XASA	0.53	-0.17	-0.33	1.00	-0.69	-0.16	0.86	0.93	0.58	0.79	-0.59	0.58
6NTI	-0.64	0.34	0.66	-0.69	1.00	0.20	-0.66	-0.67	-0.36	-0.77	0.25	-0.93
6NTA	0.43	0.74	0.58	-0.16	0.20	1.00	-0.47	-0.31	0.46	0.20	0.16	-0.13
6SFT	0.16	-0.62	-0.66	0.86	-0.66	-0.47	1.00	0.96	0.26	0.55	-0.39	0.54
6PRO	0.29	-0.41	-0.46	0.93	-0.67	-0.31	0.96	1.00	0.51	0.63	-0.53	0.50
6CHL	0.67	0.50	0.33	0.58	-0.36	0.46	0.26	0.51	1.00	0.64	-0.48	0.21
6CVT	0.88	0.12	-0.31	0.79	-0.77	0.20	0.55	0.63	0.64	1.00	-0.16	0.81
6COF	-0.14	-0.09	-0.22	-0.59	0.25	0.16	-0.39	-0.53	-0.48	-0.16	1.00	0.03
6ALC	0.70	-0.28	-0.71	0.58	-0.93	-0.13	0.54	0.50	0.21	0.81	0.03	1.00

question becomes, "do people now living in a city represent the same sample as individuals who lived in cities 14-15 years before (the lags on smoking and diet, respectively)?" If the answer is no, and if people carry their dietary and smoking characteristics with them as they move, the most recent available data is likely to better represent long-term dietary and smoking patterns of individuals in a particular city. For these reasons, in this study, we use the available data closest to 1970 throughout. However, it may well be, for diseases with long lags such as cancer, that lagged variables are superior in any case. The real answer is, of course, to account properly for mobility -- a near impossibility when using aggregate data.

The second consideration in specifying variables for inclusion is multicollinearity. Typicall, multicollinearity problems can be avoided if the simple correlations between explanatory variables are less than 0.4 to 0.6. Tables 4.7 and 4.8 present simple correlation matrices for 1965 diet and air quality variables, respectively.

Table 4.7 shows that a very high level of collinearity is probably present among dietary variables. It appears so high, in fact, that the problem becomes one of finding a set of dietary variables which is sufficiently non-collinear to allow reasonable estimation of individual coefficients. Perhaps the broadest indicators of dietary patterns are the total protein (XPRO), total fat (XFAT) and total carbohydrate (XCAR) variables. Protein and fat will tend to indicate high consumption of meat and nuts, while high carbohydrate consumption will indicate consumption of grains, fruits, vegetables, and refined sugars. However, total carbohydrate and total fat have a correlation coefficient of 0.85, too high to allow inclusion of both variables in an estimated equation. However, if we replace total fats (XFAT) with animal fat (6SFT), a good proxy for consumption of saturated fats, the correlation between fat (now animal fat) and carbohydrates drops to 0.66, which although still high, will likely cause less difficulty. Thus, we include these three diet variables as broad indicators of dietary patterns where, however, it must be clearly recognized that the estimated coefficients on these variables may well include the effect partially or totally of a number of other highly collinear dietary variables. For example, total protein (XPRO) has a correlation with cirrhosis, one might justifiably doubt that a causal relationship exists between protein and cirrhosis as opposed to one between alcohol and cirrhosis.

Table 4.8 suggests that multicollinearity may well be a problem within the air quality data set as well. Given previous research (see, for example, Lave and Seskin, 1977), the air quality measures of most concern are those for NO_2, SO_2, sulfate, and particulates, so we focus on these variables here. However, our measures of SO_2 and sulfate for 1970, the year of the mortality data, are highly collinear -- SO70 and SU70 have a simple correlation of 0.74 -- so any separation of their relative importance is likely impossible. As a result, we use SO_2 (SO70) as a proxy variable for both pollutants. Note also that among the included air pollution variables, NO_2 (NO60), particulates (PA70), and SO_2 (SO70), collinearity problems may exist with respect to ammonium, carbon monoxide and lead (some correlations greater than or equal to 0.4). Since we exclude these variables here (as

Table 4.8

Simple Correlation Matrix for Air Quality Variables

	AM66	NI66	SU66	NO69	FA66	FA70	SO70	NI70	NI90	LEAD	CO74	SU70	SU90
AM66	1.00	-0.16	0.78	0.28	0.51	0.32	0.52	-0.15	-0.08	0.04	0.14	0.66	0.64
NI66	-0.16	1.00	0.07	0.36	0.24	0.08	0.04	0.48	0.44	0.51	0.19	0.07	-0.01
SU66	0.78	0.07	1.00	0.34	0.68	0.43	0.72	0.08	0.11	0.07	0.28	0.86	0.73
NO69	0.28	0.36	0.34	1.00	0.30	0.09	0.35	0.31	0.33	0.40	0.21	0.37	0.21
FA66	0.51	0.24	0.68	0.30	1.00	0.69	0.50	0.11	0.12	0.19	0.36	0.53	0.53
FA70	0.32	0.08	0.43	0.09	0.69	1.00	0.25	0.17	0.11	-0.00	0.36	0.37	0.31
SO70	0.52	0.04	0.72	0.35	0.50	0.25	1.00	0.20	0.28	0.10	0.50	0.74	0.56
NI70	-0.15	0.48	0.08	0.31	0.11	0.17	0.20	1.00	0.92	0.18	0.24	0.19	0.04
NI90	-0.08	0.44	0.11	0.33	0.12	0.11	0.28	0.92	1.00	0.20	0.24	0.22	0.14
LEAD	0.04	0.51	0.07	0.40	0.19	-0.00	0.10	0.18	0.20	1.00	0.20	0.00	-0.08
CO74	0.14	0.19	0.28	0.21	0.36	0.36	0.50	0.24	0.24	0.20	1.00	0.34	0.25
SU70	0.66	0.07	0.86	0.37	0.53	0.37	0.74	0.19	0.22	0.00	0.34	1.00	0.85
SU90	0.64	-0.01	0.73	0.21	0.53	0.31	0.56	0.04	0.14	-0.08	0.25	0.85	1.00

Table 4.9

Simple Correlation Matrix for Included Variables

```
      NONW MAGE IN69 EDUC DENS COLD CI68 XPRO XCAR 6SFT NO69 SO70 PA70
NONW  1.00 0.02-0.26-0.22 0.40-0.04 0.02 0.02 0.23 0.02 0.22 0.29 0.03
MAGE  0.02 1.00 0.08-0.14-0.30 0.01-0.05 0.22-0.24 0.32 0.21 0.25 0.13
IN69 -0.26 0.08 1.00 0.49-0.32 0.02 0.05 0.20-0.23-0.05 0.12-0.12-0.25
EDUC -0.22-0.14 0.49 1.00-0.22 0.04-0.28 0.26-0.35-0.13-0.16-0.34-0.24
DENS  0.40-0.30-0.32-0.22 1.00-0.38-0.08 0.03 0.39-0.26 0.08-0.03-0.07
COLD -0.04 0.01 0.02 0.04-0.38 1.00 0.33-0.60-0.38 0.34-0.03 0.35 0.34
CI68  0.02-0.05 0.05-0.28-0.08 0.33 1.00-0.24-0.01 0.09 0.17 0.23-0.08
XPRO  0.02 0.22 0.20 0.26 0.03-0.60-0.24 1.00-0.01 0.16 0.11-0.08-0.31
XCAR  0.23-0.24-0.23-0.35 0.39-0.38-0.01-0.01 1.00-0.66-0.18-0.33-0.03
6SFT  0.02 0.32-0.05-0.13-0.26 0.34 0.09 0.16-0.66 1.00 0.16 0.59 0.12
NO69  0.22 0.21 0.12-0.16 0.08-0.03 0.17 0.11-0.18 0.16 1.00 0.35 0.09
SO70  0.29 0.25-0.12-0.34-0.03 0.35 0.23-0.08-0.33 0.59 0.35 1.00 0.25
PA70  0.03 0.13-0.25-0.24-0.07 0.34-0.08-0.31-0.03 0.12 0.09 0.25 1.00
```

Table 4.10
Complete Listing of Data Used

			MO70	VA70	HA70	FN70	EM70
			11.283	1.566	4.216	0.375	0.178
			2.161	0.395	1.078	0.114	0.059
1	BIRMG	AL	14.008	2.758	3.925	0.349	0.216
2	MONTG	AL	10.136	1.619	2.954	0.195	0.127
3	PHOEN	AZ	8.598	1.040	2.884	0.359	0.279
4	TUCSN	AZ	10.151	1.327	3.347	0.373	0.422
5	LROCK	AK	11.480	1.927	4.155	0.271	0.098
6	LONGB	CA	12.074	1.782	5.106	0.332	0.212
7	OAKLD	CA	11.792	1.628	4.500	0.281	0.195
8	SANDI	CA	7.658	1.214	2.728	0.165	0.148
9	SFRAN	CA	13.112	1.808	4.618	0.440	0.187
10	HARTF	CT	11.246	1.329	4.025	0.40[illegible]	0.171
11	NEWHA	CT	11.111	1.314	4.597	0.39[illegible]	0.116
12	WATER	CT	11.219	1.814	4.897	0.324	0.176
13	WASHG	DC	11.700	1.249	3.845	0.590	0.132
14	ATLAN	GA	11.809	1.636	3.596	0.515	0.157
15	CHICA	IL	12.317	1.363	5.555	0.413	0.117
16	ECHIG	IN	12.239	1.894	4.300	0.447	0.128
17	INDIA	IN	8.171	1.280	2.889	0.247	0.117
18	SBEND	IN	10.814	1.847	3.735	0.334	0.127
19	DESMO	IO	11.117	1.685	4.517	0.314	0.239
20	DUBUQ	IO	10.945	1.685	4.718	0.658	0.128
21	TOPEK	KA	9.247	1.528	3.720	0.264	0.168
22	WICHI	KA	8.393	1.276	3.052	0.253	0.199
23	LEXNG	KY	9.377	1.443	3.588	0.342	0.065
24	LOUIS	KY	13.539	2.150	5.057	0.387	0.293
25	NEWOR	LA	12.408	1.523	4.829	0.361	0.158
26	BALTI	MD	12.614	1.242	4.921	0.350	0.152
27	DETRO	MI	12.184	1.375	4.615	0.436	0.197
28	GDRAP	MI	11.986	1.771	4.604	0.385	0.182
29	MINNE	MN	12.661	2.125	4.784	0.414	0.193
30	KCITY	MO	11.513	1.668	3.431	0.333	0.199
31	STLOU	MO	15.602	2.134	6.130	0.546	0.222
32	OMAHA	NB	9.513	1.417	3.501	0.328	0.204
33	CAMDN	NJ	12.696	1.307	5.383	0.439	0.224
34	JCITY	NJ	12.946	1.305	5.703	0.526	0.196
35	NEWRK	NJ	11.616	1.043	4.712	0.382	0.133
36	ALBUQ	NM	8.230	0.993	2.162	0.308	0.205
37	NYORK	NY	11.526	1.14[illegible]	4.710	0.514	0.119
38	AKRON	OH	10.609	1.503	3.805	0.240	0.200
39	CINNC	OH	13.385	1.686	5.299	0.522	0.214
40	COLUM	OH	9.489	1.416	3.378	0.302	0.187
41	DAYTN	OH	11.925	1.646	4.425	0.427	0.250
42	TOLED	OH	11.141	1.701	4.343	0.380	0.221
43	YTOWN	OH	13.063	1.953	4.929	0.486	0.186
44	TULSA	OK	8.561	1.345	2.943	0.244	0.148
45	PLAND	OR	13.517	2.287	4.843	0.384	0.243
46	PHILA	PN	12.553	1.376	4.486	0.405	0.148
47	PITTS	PN	14.062	1.873	6.127	0.498	0.115
48	REDDG	PN	15.198	2.407	6.412	0.297	0.160
49	PROVI	RI	13.727	1.663	5.731	0.379	0.156
50	CHATT	TN	15.141	2.242	5.895	0.554	0.202
51	MEMPH	TN	9.358	1.536	3.222	0.269	0.159
52	DALLS	TX	8.423	1.108	2.834	0.261	0.159
53	HOUST	TX	7.803	1.001	2.614	0.266	0.119
54	PASSA	TX	5.018	0.526	1.613	0.202	0.112
55	SANAN	TX	8.131	1.217	2.606	0.266	0.125
56	SALTC	UT	11.246	1.478	3.684	0.273	0.262
57	NFOLK	VA	8.008	1.085	2.939	0.260	0.094
58	RICHM	VA	12.503	1.899	4.627	0.501	0.140
59	SEATT	WA	12.047	1.850	4.372	0.377	0.213
60	CHARL	WV	14.321	1.510	6.028	0.727	0.280

Table 4.10
(continued)

	CI70	NE70	C\BZ	I\BZ	CA70
	0.238	0.058	0.473	1.294	1.958
	0.106	0.027	0.105	0.333	0.402
1 BIRMG AL	0.116	0.116	0.465	1.427	2.343
2 MONTG AL	0.112	0.112	0.243	1.335	1.732
3 PHOEN AZ	0.205	0.029	0.246	0.975	1.499
4 TUCSN AZ	0.198	0.027	0.688	0.995	1.719
5 LROCK AK	0.188	0.045	0.478	1.299	1.942
6 LONGB CA	0.290	0.028	0.482	1.045	2.049
7 OAKLD CA	0.358	0.047	0.401	1.250	2.090
8 SANDI CA	0.191	0.030	0.467	1.058	1.316
9 SFRAN CA	0.646	0.049	0.360	0.934	2.493
10 HARTF CT	0.329	0.032	0.506	1.519	1.987
11 NEWHA CT	0.211	0.022	0.513	1.758	2.004
12 WATER CT	0.333	0.037	0.396	1.189	1.999
13 WASHG DC	0.489	0.081	0.401	1.937	2.000
14 ATLAN GA	0.213	0.131	0.355	1.539	1.789
15 CHICA IL	0.289	0.062	0.430	1.612	2.038
16 ECHIG IN	0.298	0.106	0.561	1.794	2.022
17 INDIA IN	0.140	0.058	0.441	1.254	1.573
18 SBEND IN	0.159	0.080	0.549	1.450	2.142
19 DESMO IO	0.130	0.020	0.619	1.381	1.919
20 DUBUQ IO	0.096	0.064	0.413	1.570	1.894
21 TOPEK KA	0.104	0.024	0.392	0.940	1.568
22 WICHI KA	0.116	0.036	0.506	1.465	1.454
23 LEXNG KY	0.176	0.037	0.359	0.934	1.517
24 LOUIS KY	0.288	0.055	0.436	1.237	2.304
25 NEWOR LA	0.189	0.091	0.609	1.497	2.158
26 BALTI MD	0.394	0.063	0.456	1.417	2.228
27 DETRO MI	0.413	0.067	0.470	1.320	2.126
28 GDRAP MI	0.187	0.035	0.579	1.183	2.155
29 MINNE MN	0.214	0.028	0.544	1.258	2.281
30 KCITY MO	0.187	0.049	0.384	1.318	1.871
31 STLOU MO	0.268	0.090	0.475	1.449	2.562
32 OMAHA NB	0.184	0.043	0.486	1.362	1.776
33 CAMDN NJ	0.419	0.068	0.319	1.716	2.048
34 JCITY NJ	0.426	0.111	0.412	1.687	2.241
35 NEWRK NJ	0.290	0.092	0.489	1.997	1.663
36 ALBUQ NM	0.254	0.025	0.518	1.132	1.276
37 NYORK NY	0.352	0.055	0.424	1.245	2.151
38 AKRON OH	0.160	0.051	0.430	1.291	1.968
39 CINNC OH	0.190	0.053	0.474	1.040	2.510
40 COLUM OH	0.143	0.041	0.456	0.946	1.740
41 DAYTN OH	0.177	0.078	0.506	1.224	2.048
42 TOLED OH	0.190	0.044	0.506	0.127	1.964
43 YTOWN OH	0.243	0.086	0.865	1.259	2.389
44 TULSA OK	0.130	0.036	0.491	1.188	1.586
45 PLAND OR	0.314	0.044	0.588	0.874	2.425
46 PHILA PN	0.219	0.066	0.535	1.713	2.170
47 PITTS PN	0.277	0.071	0.490	1.510	2.426
48 REDDG PN	0.251	0.023	0.676	1.149	2.818
49 PROVI RI	0.262	0.084	0.600	1.327	2.690
50 CHATT TN	0.218	0.076	0.556	2.105	2.469
51 MEMPH TN	0.154	0.056	0.572	1.143	1.644
52 DALLS TX	0.165	0.034	0.512	1.297	1.397
53 HOUST TX	0.166	0.039	0.436	1.264	1.311
54 PASSA TX	0.101	0.011	0.526	0.789	0.941
55 SANAN TX	0.194	0.064	0.377	1.182	1.373
56 SALTC UT	0.222	0.080	0.296	0.765	1.558
57 NFOLK VA	0.146	0.052	0.411	1.447	1.322
58 RICHM VA	0.248	0.072	0.377	1.684	1.959
59 SEATT WA	0.266	0.053	0.416	1.075	2.174
60 CHARL WV	0.392	0.098	0.421	0.843	2.671

Table 4.10
(continued)

			MAGE	CI68	MDOC
			28.820	165.798	162.807
			2.737	23.247	54.226
1	BIRMG	AL	29.700	136.604	142.000
2	MONTG	AL	27.700	136.604	80.400
3	PHOEN	AZ	27.500	128.693	148.500
4	TUCSN	AZ	26.800	128.693	166.600
5	LROCK	AK	29.600	150.591	229.800
6	LONGB	CA	32.700	157.477	186.800
7	OAKLD	CA	31.400	157.477	239.600
8	SANDI	CA	25.800	157.477	149.400
9	SFRAN	CA	34.200	157.477	239.600
10	HARTF	CT	27.800	201.869	195.100
11	NEWHA	CT	27.800	201.869	327.300
12	WATER	CT	31.300	201.869	377.300
13	WASHG	DC	28.400	137.765	203.400
14	ATLAN	GA	27.200	159.601	159.500
15	CHICA	IL	29.600	189.522	161.400
16	ECHIG	IN	27.700	191.140	68.700
17	INDIA	IN	27.100	191.140	141.400
18	SBEND	IN	29.600	191.140	120.300
19	DESMO	IO	28.600	158.907	175.500
20	DUBUQ	IO	25.000	158.907	105.300
21	TOPEK	KA	28.000	146.046	127.500
22	WICHI	KA	27.000	146.046	111.500
23	LEXNG	KY	24.600	203.395	98.900
24	LOUIS	KY	30.100	203.395	159.500
25	NEWOR	LA	27.900	182.361	209.100
26	BALTI	MD	28.700	166.384	200.400
27	DETRO	MI	29.300	202.620	152.700
28	GDRAP	MI	27.100	202.620	125.900
29	MINNE	MN	29.100	166.797	156.800
30	KCITY	MO	29.500	166.797	156.600
31	STLOU	MO	31.400	164.695	144.800
32	OMAHA	NB	26.700	152.560	160.100
33	CAMDN	NJ	27.500	148.636	158.200
34	JCITY	NJ	30.700	148.636	101.900
35	NEWRK	NJ	25.900	148.636	195.400
36	ALBUQ	NM	25.100	157.395	188.100
37	NYORK	NY	32.400	152.273	250.200
38	AKRON	OH	28.500	183.177	143.400
39	CINNC	OH	28.800	183.177	157.000
40	COLUM	OH	25.400	183.177	186.300
41	DAYTN	OH	27.800	183.177	113.500
42	TOLED	OH	28.700	183.177	128.600
43	YTOWN	OH	31.600	183.177	128.600
44	TULSA	OK	28.700	111.392	143.400
45	PLAND	OR	32.600	170.820	186.200
46	PHILA	PN	30.900	147.692	207.000
47	PITTS	PN	33.400	147.692	144.300
48	REDDG	PN	36.900	147.692	112.400
49	PROVI	RI	32.100	191.745	157.000
50	CHATT	TN	31.100	160.097	120.700
51	MEMPH	TN	26.100	160.097	167.900
52	DALLS	TX	27.200	171.593	154.800
53	HOUST	TX	26.000	171.593	144.500
54	PASSA	TX	24.400	171.593	144.500
55	SANAN	TX	24.800	171.593	119.600
56	SALTC	UT	27.700	100.734	193.600
57	NFOLK	VA	24.000	172.618	93.600
58	RICHM	VA	29.500	172.618	231.100
59	SEATT	WA	31.900	143.939	185.300
60	CHARL	WV	34.600	181.301	89.700

Table 4.10
(continued)

	DENS	NONW	IN69	EDUC	COLD
	0.022	0.226	10762.895	55.298	86.900
	0.013	0.154	1059.562	7.439	47.654
1 BIRMG AL	0.037	0.422	8692.000	45.400	48.000
2 MONTG AL	0.037	0.336	9933.000	51.600	29.000
3 PHOEN AZ	0.027	0.067	11329.000	60.100	13.000
4 TUCSN AZ	0.032	0.052	9922.000	63.100	22.000
5 LROCK AK	0.017	0.251	10438.000	56.500	51.000
6 LONGB CA	0.013	0.082	11804.000	62.000	1.000
7 OAKLD CA	0.023	0.407	11279.000	66.100	7.000
8 SANDI CA	0.019	0.111	11664.000	65.300	0.0
9 SFRAN CA	0.027	0.286	12507.000	66.100	11.000
10 HARTF CT	0.029	0.292	10011.000	59.100	147.000
11 NEWHA CT	0.015	0.274	10444.000	56.800	102.000
12 WATER CT	0.015	0.105	11500.000	49.900	102.000
13 WASHG DC	0.048	0.723	12189.000	68.500	114.000
14 ATLAN GA	0.034	0.516	10656.000	53.400	45.000
15 CHICA IL	0.026	0.344	13527.000	53.900	132.000
16 ECHIG IN	0.047	0.284	10068.000	53.600	132.000
17 INDIA IN	0.017	0.184	12260.000	56.000	122.000
18 SBEND IN	0.010	0.147	11431.000	54.200	131.000
19 DESMO IO	0.010	0.062	11350.000	68.000	151.000
20 DUBUQ IO	0.018	0.003	11274.000	54.600	152.000
21 TOPEK KA	0.011	0.096	10830.000	64.800	122.000
22 WICHI KA	0.013	0.107	10940.000	63.200	122.000
23 LEXNG KY	0.026	0.174	10033.000	60.100	94.000
24 LOUIS KY	0.019	0.241	9980.000	46.900	91.000
25 NEWOR LA	0.055	0.045	10246.000	45.800	13.000
26 BALTI MD	0.017	0.470	10035.000	44.600	90.000
27 DETRO MI	0.016	0.445	11015.000	52.100	126.000
28 GDRAP MI	0.007	0.120	11242.000	54.000	157.000
29 MINNE MN	0.010	0.064	11127.000	66.100	167.000
30 KCITY MO	0.017	0.228	11306.000	60.100	106.000
31 STLOU MO	0.044	0.413	9268.000	48.000	118.000
32 OMAHA NB	0.013	0.106	11605.000	62.700	144.000
33 CAMDN NJ	0.018	0.402	8627.000	50.600	89.000
34 JCITY NJ	0.026	0.222	10285.000	36.300	79.000
35 NEWRK NJ	0.040	0.560	8637.000	55.100	84.000
36 ALBUQ NM	0.022	0.043	10926.000	66.200	122.000
37 NYORK NY	0.029	0.234	11632.898	51.800	79.000
38 AKRON OH	0.007	0.178	11152.000	55.600	133.000
39 CINNC OH	0.028	0.281	10435.000	48.400	100.000
40 COLUM OH	0.010	0.190	10848.000	60.700	114.000
41 DAYTN OH	0.015	0.309	10329.000	56.200	125.000
42 TOLED OH	0.001	0.143	11590.000	51.700	157.000
43 YTOWN OH	0.011	0.258	9928.000	52.100	150.000
44 TULSA OK	0.010	0.134	11642.000	58.200	82.000
45 PLAND OR	0.008	0.078	11377.000	62.900	42.000
46 PHILA PN	0.013	0.344	10431.000	50.600	89.000
47 PITTS PN	0.013	0.207	10536.000	53.400	124.000
48 REDDG PN	0.006	0.068	9695.000	43.300	89.000
49 PROVI RI	0.010	0.100	10208.000	45.900	128.000
50 CHATT TN	0.026	0.360	8336.000	47.600	70.000
51 MEMPH TN	0.049	0.392	10104.000	49.200	61.000
52 DALLS TX	0.028	0.258	12474.000	54.800	34.000
53 HOUST TX	0.032	0.266	11737.000	51.700	21.000
54 PASSA TX	0.018	0.005	11822.000	51.700	21.000
55 SANAN TX	0.062	0.086	9027.000	46.800	23.000
56 SALTC UT	0.015	0.032	10812.000	68.500	110.000
57 NFOLK VA	0.021	0.302	9236.000	48.300	31.000
58 RICHM VA	0.018	0.424	10620.000	47.100	68.000
59 SEATT WA	0.010	0.126	12557.000	67.800	32.000
60 CHARL WV	0.012	0.108	10865.000	52.800	95.000

Table 4.10
(continued)

			6SFT	XCAR	APRO
			16314.051	123491.313	39844.547
			975.743	3618.011	706.747
1	BIRMG	AL	15079.500	129119.000	39517.699
2	MONTG	AL	15030.398	128765.000	39264.801
3	PHOEN	AZ	16116.699	120959.000	40696.199
4	TUCSN	AZ	16054.102	120957.000	40395.500
5	LROCK	AK	15186.500	129263.000	39630.301
6	LONGB	CA	16079.102	120519.000	41050.398
7	OAKLD	CA	16030.898	120345.000	41306.898
8	SANDI	CA	16090.398	120865.000	40788.801
9	SFRAN	CA	16030.898	120345.000	41306.898
10	HARTF	CT	18240.500	121166.000	40740.000
11	NEWHA	CT	18118.801	120707.000	40348.500
12	WATER	CT	18255.801	120882.000	40413.301
13	WASHG	DC	16219.500	126710.000	40232.000
14	ATLAN	GA	15737.398	128113.000	40143.199
15	CHICA	IL	16404.699	121525.000	39337.301
16	ECHIG	IN	16295.000	121512.000	39144.000
17	INDIA	IN	16261.102	121447.000	39100.199
18	SBEND	IN	16198.398	121453.000	39005.500
19	DESMO	IO	16244.699	121511.000	39096.000
20	DUBUQ	IO	16153.102	121495.000	38978.102
21	TOPEK	KA	16070.801	121291.000	38877.500
22	WICHI	KA	16040.102	121060.000	38827.102
23	LEXNG	KY	15555.602	128726.000	39981.301
24	LOUIS	KY	15563.398	128982.000	40212.000
25	NEWOR	LA	15119.199	128678.000	39532.602
26	BALTI	MD	15724.500	128374.000	40248.398
27	DETRO	MI	16419.898	121501.000	39346.500
28	GDRAP	MI	16266.801	121560.000	39095.500
29	MINNE	MN	16431.801	121751.000	39327.000
30	KCITY	MO	16217.199	121345.000	39047.199
31	STLOU	MO	16175.398	121333.000	39030.199
32	OMAHA	NB	16177.398	121345.000	38997.801
33	CAMDN	NJ	17976.199	119345.000	39194.500
34	JCITY	NJ	18157.500	120490.000	39916.699
35	NEWRK	NJ	18148.801	121000.000	40602.398
36	ALBUQ	NM	16052.102	120684.000	40456.801
37	NYORK	NY	18028.102	120486.000	40287.301
38	AKRON	OH	16307.500	121535.000	39152.102
39	CINNC	OH	16160.301	121335.000	39006.602
40	COLUM	OH	16194.898	121321.000	39035.602
41	DAYTN	OH	16349.199	121493.000	39215.699
42	TOLED	OH	16273.801	121546.000	39135.398
43	YTOWN	OH	16232.602	121441.000	39032.500
44	TULSA	OK	15467.898	129016.000	39958.602
45	PLAND	OR	16006.199	121115.000	40881.301
46	PHILA	PN	18187.398	120737.000	40271.699
47	PITTS	PN	18287.898	120748.000	39978.602
48	REDDG	PN	18399.699	121050.000	40029.699
49	PROVI	RI	18237.699	120540.000	39969.000
50	CHATT	TN	15181.398	129289.000	39708.500
51	MEMPH	TN	15129.301	128846.000	39551.301
52	DALLS	TX	15717.000	128359.000	40155.199
53	HOUST	TX	15626.500	128489.000	40134.500
54	PASSA	TX	15947.398	128755.000	40795.199
55	SANAN	TX	15080.699	129045.000	39443.699
56	SALTC	UT	16013.500	121539.000	40791.102
57	NFOLK	VA	15177.000	128966.000	39734.602
58	RICHM	VA	15640.699	128727.000	40184.699
59	SEATT	WA	16040.199	120738.000	41333.699
60	CHARL	WV	15233.301	129243.000	39718.398

Table 4.10
(continued)

			NO69	PA70	SO70
			0.076	102.300	26.917
			0.034	30.107	22.197
1	BIRMG	AL	0.093	138.000	10.000
2	MONTG	AL	0.016	91.000	7.000
3	PHOEN	AZ	0.089	117.000	9.000
4	TUCSN	AZ	0.028	104.000	7.000
5	LROCK	AK	0.012	71.000	8.000
6	LONGB	CA	0.182	102.000	35.000
7	OAKLD	CA	0.053	69.000	10.000
8	SANDI	CA	0.106	88.000	10.000
9	SFRAN	CA	0.095	52.000	9.000
10	HARTF	CT	0.077	67.000	57.000
11	NEWHA	CT	0.072	100.000	40.000
12	WATER	CT	0.037	95.000	17.000
13	WASHG	DC	0.069	99.000	28.000
14	ATLAN	GA	0.096	88.000	20.000
15	CHICA	IL	0.160	58.000	73.000
16	ECHIG	IN	0.086	189.000	87.000
17	INDIA	IN	0.079	96.000	51.000
18	SBEND	IN	0.031	99.000	10.000
19	DESMO	IO	0.033	134.000	11.000
20	DUBUQ	IO	0.072	194.000	16.000
21	TOPEK	KA	0.023	97.000	8.000
22	WICHI	KA	0.064	95.000	6.000
23	LEXNG	KY	0.062	71.000	12.000
24	LOUIS	KY	0.096	117.000	23.000
25	NEWOR	LA	0.061	79.000	7.000
26	BALTI	MD	0.099	136.000	54.000
27	DETRO	MI	0.116	128.000	38.000
28	GDRAP	MI	0.090	77.000	13.000
29	MINNE	MN	0.076	88.000	38.000
30	KCITY	MO	0.045	84.000	6.000
31	STLOU	MO	0.135	120.000	42.000
32	OMAHA	NB	0.075	132.000	12.000
33	CAMDN	NJ	0.128	114.000	69.000
34	JCITY	NJ	0.065	108.000	75.000
35	NEWRK	NJ	0.092	111.000	37.000
36	ALBUQ	NM	0.048	102.000	6.000
37	NYORK	NY	0.142	109.000	45.000
38	AKRON	OH	0.038	68.000	70.000
39	CINNC	OH	0.099	95.000	18.000
40	COLUM	OH	0.087	97.000	22.000
41	DAYTN	OH	0.057	102.000	25.000
42	TOLED	OH	0.096	111.000	13.000
43	YTOWN	OH	0.083	134.000	30.000
44	TULSA	OK	0.033	110.000	8.000
45	PLAND	OR	0.058	60.000	14.000
46	PHILA	PN	0.024	144.000	78.000
47	PITTS	PN	0.113	137.000	57.000
48	REDDG	PN	0.081	121.000	30.000
49	PROVI	RI	0.087	93.000	67.000
50	CHATT	TN	0.042	127.000	18.000
51	MEMPH	TN	0.078	86.000	17.000
52	DALLS	TX	0.074	105.000	7.000
53	HOUST	TX	0.105	90.000	10.000
54	PASSA	TX	0.050	80.000	9.000
55	SANAN	TX	0.065	58.000	8.000
56	SALTC	UT	0.060	88.000	9.000
57	NFOLK	VA	0.077	83.000	26.000
58	RICHM	VA	0.088	89.000	24.000
59	SEATT	WA	0.095	58.000	22.000
60	CHARL	WV	0.092	183.000	27.000

have most previous investigators) some unknown contribution from these pollutants may be included in estimated coefficients on NO_2, SO_2, and particulates.

A complete correlation matrix for our choice of included variables is presented in Table 4.9 with the exception of the doctors per capita variable for which we use two-stage least squares. Several interesting collinearity issues are apparent here as well, not necessarily as a multicollinearity problem since the highest simple correlation is 0.66 within the data set, but rather as an indication of problems of exclusion of variables in previous studies. For example, SO_2 (SO70) shows a correlation of 0.59 with animal fat consumption (6SFT) and a correlation of 0.23 with cigarette consumption. In other words, air pollution as measured by SO_2 might not be orthogonal with respect to diet and smoking. The implication is that diet and smoking probably must be included to obtain unbiased estimates of the effect of SO_2 on human health.

The final data set used in the analysis is presented is its entirely in Table 4.10. [Qualified investigators wishing to know about the additional data collected may obtain a complete listing by contacting the authors].

4.5 Empirical Analysis

The first step in estimating the model of human health we have specified is to attempt to account for human adjustments to disease in the form of medical care (doctors per capita). Thus, we first estimate a reduced form equation which has as the dependent variable doctors per capita and as independent, exogenous explanatory variables: median age, % non-white, density, cold, per capita consumption of cigarettes, protein, carbohydrates, and animal fat (all exogenous variables from the dose response function); as well as per capita income and education level (exogenous variables in the determination of the demand for doctors). This estimated equation is shown in Table 4.12 below.

Note that we have chosen linear specifications throughout the analysis. The linear form has several advantages. First, the entire modeling framework can be interpreted as providing a set of first-order approximations of the slopes (the effects) of the variables in the model. As a linear approximate system, the estimated effects, if unbiased, only hold true for the neighborhood of the estimate -- that is for values of the variables near the means of our sample. Since we do not know the precise form (presumably nonlinear) which our functions take, we are at least not implying more about our knowledge of effects than the data can support. However, if significant nonlinearities do exist over the range of the data in our sample (and some of the estimated effects seem to imply this), then we have introduced a specification error by choosing linear estimates.

The second step in our analysis, then, is to generate, from the estimated reduced form equation and from data on exogenous variables in the model, an estimated doctors per capita variable (DOCH) to replace actual data on doctors per capita(MDOC) in the estimation of specific dose-response functions. Table 4.11 summarizes our results, while Tables 4.13-4.22 give

Table 4.11

Summary of Two-Stage Linear Estimates of Factors in Human Mortality
Hypotheses not Rejected at the 97.5% Confidence Level
(One-tailed t-test, t $\geq$ 2.0)

Variable (Sign of Hypothetical Effect)		Total Mortality Rate	Vascular Disease	Heart Disease	Pneumonia and Influenza	Emphysema and Bronchitis	Cirrhosis	Kidney Disease	Congenital Birth Defects	Early Infant Diseases	Cancer
Doctors/Capita*	(-)	-	-	-		-		-			-
Median Age	(+)	+	+	+	+		+	+			+
% Nonwhite	(+)	+		+			+	+		+	+
Cigarettes	(+)	+	+	+							+
Room Density	(+)	+			+		+	+			
Cold	(+)	+			+						+
Animal Fat	(+)			+							
Protein	(+)	+				+					+
Carbohydrates	(?)					-					
NO_2	(+)										
SO_2	(+)									+	
Particulates	(+)				+						
R^2		.82	.60	.77	.54	.39	.64	.54	.22	.55	.86

* Two-stage estimator employed.

individual linear estimated dose-response functions where the first equation reported excludes the air quality variables and the second (for comparison) includes these variables. Note that the basic model actually excludes air quality. This reflects the approach taken; to first develop a model of human health based on variables which are hypothesized to have large effects -- age, medical care, smoking, diet, etc. Once a basic model of satisfactory explanatory power is specified, it then may be appropriate to test for variables such as air pollution which are hypothesized to have small effects.

We choose a 97.5% confidence level ($t \geq 2.0$) for the entire analysis. It should be noted that if actual doctors per capita are employed, the variable is highly nonsignificant both when used in total mortality or when used in any of the component diseases. However, as indicated in our summary of results shown in Table 4.11 the estimated two-stage doctors per capita variable is highly significant and has a uniform negative effect on mortality rates in total mortality, vascular disease, heart disease, emphysema and bronchitis, kidney disease, and cancer. We conclude, then, from a properly structured hypothesis test, that we cannot reject the hypothesis that medical care has a highly important effect on human mortality.

Both the median age and % nonwhite variables are widely significant across the estimated equations and show up with uniformly positive effects on mortality rates.

Cigarette consumption shows significant positive partial correlations with total mortality, vascular disease, heart disease and cancer, while room density and cold both show significant positive partial correlations with total mortality and pneumonia and influenza. Room density also shows significant positive correlations for cirrhosis and kidney disease.

The dietary variables show significance in total mortality, heart disease, and cancer -- correlations between heart disease and saturated fats and between cancer and meat consumption (note the positive association for protein) have long been recognized -- as well as in emphysema and bronchitis.

Again, however, it must be stressed, especially for the dietary variables, that collinearity abounds and the estimated effect may really not be related to the specified variable but to a highly collinear one. Similarly, an estimated effect may include the sum of the effects of several collinear variables. In other words, causality is not established by correlation. However, we cannot reject the hypotheses that doctors, cigarettes, and diet are all highly important in determining human mortality rates. Unfortunately, these variables have typically been excluded from previous aggregate epidemiological studies [Schwing, et. al. (1974) do, however, include a smoking variable].

Turning to the air quality variables, only two significant partial correlations appear -- between particulates and the pneumonia and influenza variable, and between SO_2 and the early infant disease variable. The latter of these effects is consistent with the work of Lave and Seskin (1977).

Table 4.12

Reduced Form Equation

```
MDOC  55 -- 10 11 13 14 21 24 54 25 27 41

     VAR                  B                  T

   X10-NONW            50.447             1.1020
   X11-MAGE            1.3513             0.50439
   X13-IN69          0.61649D-02          0.86708
   X14-EDUC            1.9399             1.3418
   X21-DENS            161.53             0.26107
   X24-COLD          -0.12806            -0.53850
   X54-CI68           0.45771             1.4919
   X25-XPRO          0.22302D-01          1.4136
   X27-XCAR          0.22804D-02          0.72804
   X41-6SFT          0.24002D-01          1.9402
   CONSTANT          -1691.3             -2.7124

R-SQUARE= 0.3877
SSR=           0.1062D+06     DF= 49
```

Table 4.13

Total Mortality

```
 MO70   1 -- 61 10 11 21 24 54 25 27 41

      VAR                    B                      T

     X61-DOCH           -0.53031D-01             -4.7501
     X10-NONW             5.6092                  5.0448
     X11-MAGE            0.66172                  12.360
     X21-DENS             31.910                  2.4494
     X24-COLD            0.14370D-01              3.1093
     X54-CI68            0.21896D-01              3.0758
     X25-XPRO            0.19325D-02              3.7525
     X27-XCAR           -0.82907D-04             -1.5347
     X41-6SFT            0.36251D-03              1.5938
     CONSTANT            -78.675                 -3.7169

R-SQUARE= 0.8195
SSR=            49.74          DF= 50

 MO70   1 -- 61 10 11 21 24 54 25 27 41 31 57 60

      VAR                    B                      T

     X61-DOCH           -0.52797D-01             -4.3493
     X10-NONW             5.6276                  4.5620
     X11-MAGE            0.65893                  11.540
     X21-DENS             31.772                  2.3469
     X24-COLD            0.14436D-01              2.9089
     X54-CI68            0.21968D-01              2.8120
     X25-XPRO            0.19196D-02              3.5522
     X27-XCAR           -0.79431D-04             -1.3612
     X41-6SFT            0.39783D-03              1.4508
     X31-NO69             1.6457                  0.35799
     X57-SO70           -0.31302D-02             -0.34850
     X60-PA70            0.10744D-02              0.20059
     CONSTANT            -79.296                 -3.5115

R-SQUARE= 0.8205
SSR=            49.48          DF= 47
```

Table 4.14

Vascular Disease

```
VA70    2 -- 61  10  11  21  24  54  25  27  41

      VAR                   B                       T

    X61-DOCH         -0.70223D-02              -2.2186
    X10-NONW          0.37954                   1.2040
    X11-MAGE          0.10936                   7.2050
    X21-DENS           1.7016                  0.46071
    X24-COLD          0.17695D-02               1.3504
    X54-CI68          0.30001D-02               1.4865
    X25-XPRO          0.18542D-03               1.2699
    X27-XCAR         -0.13460D-04             -0.87885
    X41-6SFT         -0.84316D-04              -1.3075
    CONSTANT          -5.5672                 -0.92769

R-SQUARE= 0.5657
SSR=            3.998           DF= 50


VA70    2 -- 61  10  11  21  24  54  25  27  41  31  57  60

      VAR                   B                       T

    X61-DOCH         -0.88088D-02              -2.6671
    X10-NONW          0.63816                   1.9014
    X11-MAGE          0.11671                   7.5127
    X21-DENS           2.7540                  0.74768
    X24-COLD          0.20961D-02               1.5524
    X54-CI68          0.42167D-02               1.9838
    X25-XPRO          0.23976D-03               1.6306
    X27-XCAR         -0.17117D-04              -1.0781
    X41-6SFT         -0.18874D-04             -0.25297
    X31-NO69         -0.86198                 -0.68915
    X57-SO70         -0.38911D-02              -1.5922
    X60-PA70         -0.63071D-03             -0.43278
    CONSTANT          -8.3456                  -1.3583

R-SQUARE= 0.6022
SSR=            3.663           DF= 47
```

Table 4.15

Heart Disease

```
 HA70    3 -- 61  10  11  21  24  54  25  27  41

         VAR                  B                      T

      X61-DOCH         -0.22340D-01              -3.4204
      X10-NONW           1.7509                   2.6917
      X11-MAGE          0.29627                   9.4592
      X21-DENS           9.3338                   1.2247
      X24-COLD          0.43566D-02               1.6113
      X54-CI68          0.13129D-01               3.1525
      X25-XPRO          0.66878D-03               2.2197
      X27-XCAR         -0.17969D-04              -0.56857
      X41-6SFT          0.42380D-03               3.1849
      CONSTANT         -35.183                   -2.8412

R-SQUARE= 0.7517
SSR=             17.03          DF= 50
```

```
 HA70    3 -- 61  10  11  21  24  54  25  27  41  31  57  60

         VAR                  B                      T

      X61-DOCH         -0.19800D-01              -2.9132
      X10-NONW           1.3603                   1.9694
      X11-MAGE          0.28122                   8.7965
      X21-DENS           7.4177                  0.97861
      X24-COLD          0.43300D-02               1.5583
      X54-CI68          0.10868D-01               2.4847
      X25-XPRO          0.57165D-03               1.8893
      X27-XCAR         -0.22071D-05            -0.67551D-01
      X41-6SFT          0.39645D-03               2.5822
      X31-NO69           4.8498                   1.8842
      X57-SO70          0.82843D-03              0.16473
      X60-PA70          0.14052D-02              0.46855
      CONSTANT         -32.823                  -2.5960

R-SQUARE= 0.7737
SSR=             15.51          DF= 47
```

Table 4.16

Pneumonia and Influenza

```
PN70    4 -- 61 10 11 21 24 54 25 27 41

     VAR                    B                      T

   X61-DOCH          -0.12925D-02              -1.2098
   X10-NONW           0.18381                   1.7274
   X11-MAGE           0.19920D-01               3.8880
   X21-DENS            2.9692                   2.3816
   X24-COLD           0.13030D-02               2.9460
   X54-CI68           0.61524D-03              0.90310
   X25-XPRO           0.70654D-04               1.4336
   X27-XCAR          -0.11329D-05             -0.21914
   X41-6SFT           0.38010D-05              0.17462
   CONSTANT           -3.0478                  -1.5046

R-SQUARE= 0.4108
SSR=           0.4556           DF= 50
```

```
PN70    4 -- 61 10 11 21 24 54 25 27 41 31 57 60

     VAR                    B                      T

   X61-DOCH          -0.38115D-03             -0.37068
   X10-NONW           0.13820                   1.3226
   X11-MAGE           0.16194D-01               3.3484
   X21-DENS            2.6677                   2.3264
   X24-COLD           0.10428D-02               2.4807
   X54-CI68           0.61816D-03              0.93415
   X25-XPRO           0.46048D-04               1.0060
   X27-XCAR          -0.10027D-05             -0.20286
   X41-6SFT          -0.49261D-05             -0.21208
   X31-NO69           0.55215                   1.4180
   X57-SO70          -0.47719D-03             -0.62722
   X60-PA70           0.14272D-02               3.1456
   CONSTANT           -2.1183                  -1.1075

R-SQUARE= 0.5409
SSR=           0.3550           DF= 47
```

Table 4.17

Emphysema and Bronchitis

```
EM70   5 -- 61 10 11 21 24 54 25 27 41

     VAR                  B                    T

   X61-DOCH          -0.13851D-02          -2.3989
   X10-NONW           0.22533D-01          0.39184
   X11-MAGE           0.49451D-02           1.7859
   X21-DENS           0.92106D-01          0.13670
   X24-COLD           0.18996D-03          0.79468
   X54-CI68           0.53419D-04          0.14509
   X25-XPRO           0.65261D-04           2.4502
   X27-XCAR          -0.95199D-05          -3.4072
   X41-6SFT          -0.17700D-04          -1.5046
   CONSTANT          -0.90721             -0.82870

R-SQUARE= 0.3559
SSR=          0.1331          DF= 50

EM70   5 -- 61 10 11 21 24 54 25 27 41 31 57 60

     VAR                  B                    T

   X61-DOCH          -0.13761D-02          -2.2419
   X10-NONW           0.46333D-01          0.74278
   X11-MAGE           0.51065D-02           1.7687
   X21-DENS           0.17013              0.24853
   X24-COLD           0.14392D-03          0.57352
   X54-CI68           0.24382D-03          0.61719
   X25-XPRO           0.66687D-04           2.4404
   X27-XCAR          -0.10427D-04          -3.5335
   X41-6SFT          -0.13868D-04          -1.0001
   X31-NO69          -0.11515             -0.49533
   X57-SO70          -0.47320D-03          -1.0419
   X60-PA70           0.29115D-03           1.0750
   CONSTANT          -0.96354             -0.84382

R-SQUARE= 0.3876
SSR=          0.1265          DF= 47
```

Table 4.18

Cirrhosis

```
  CI70    6 -- 61  10  11  21  24  54  25  27  41

        VAR                    B                      T

       X61-DOCH          0.51923D-04            0.65875D-01
       X10-NONW          0.22721                  2.8943
       X11-MAGE          0.17323D-01              4.5831
       X21-DENS           2.1612                  2.3497
       X24-COLD          0.52436D-03              1.6070
       X54-CI68          0.61784D-03              1.2293
       X25-XPRO          0.73982D-04              2.0347
       X27-XCAR         -0.77437D-05             -2.0303
       X41-6SFT         -0.11197D-04             -0.69726
       CONSTANT          -2.3252                 -1.5559

 R-SQUARE= 0.6258
 SSR=             0.2479            DF= 50

  CI70    6 -- 61  10  11  21  24  54  25  27  41  31  57  60

        VAR                    B                      T

       X61-DOCH          0.31818D-03              0.37741
       X10-NONW          0.18469                  2.1558
       X11-MAGE          0.16077D-01              4.0543
       X21-DENS           1.9819                  2.1080
       X24-COLD          0.49340D-03              1.4315
       X54-CI68          0.39600D-03              0.72987
       X25-XPRO          0.65122D-04              1.7351
       X27-XCAR         -0.67906D-05             -1.6755
       X41-6SFT         -0.19734D-04             -1.0362
       X31-NO69          0.24209                  0.75825
       X57-SO70          0.51883D-03              0.83173
       X60-PA70          0.69311D-04              0.18632
       CONSTANT          -1.9446                 -1.2399

 R-SQUARE= 0.6399
 SSR=             0.2387            DF= 47
```

Table 4.19

Kidney Disease

```
 NE70    7 -- 61  10  11  21  24  54  25  27  41

       VAR                    B                       T

     X61-DOCH          -0.67302D-03             -2.9834
     X10-NONW           0.90661D-01              4.0353
     X11-MAGE           0.34110D-02              3.1531
     X21-DENS           0.75879                  2.8825
     X24-COLD           0.73289D-04             0.78479
     X54-CI68           0.10132D-03             0.70439
     X25-XPRO           0.15723D-04              1.5110
     X27-XCAR           0.47349D-07             0.43376D-01
     X41-6SFT           0.63724D-05              1.3866
     CONSTANT          -0.72766                 -1.7013

R-SQUARE= 0.5419
SSR=            0.2031D-01      DF= 50

 NE70    7 -- 61  10  11  21  24  54  25  27  41  31  57  60

       VAR                    B                       T

     X61-DOCH          -0.55265D-03             -2.3310
     X10-NONW           0.79829D-01              3.3134
     X11-MAGE           0.30851D-02              2.7665
     X21-DENS           0.71889                  2.7189
     X24-COLD           0.29075D-04             0.29997
     X54-CI68           0.80286D-04             0.52619
     X25-XPRO           0.12995D-04              1.2312
     X27-XCAR          -0.20721D-06            -0.18181
     X41-6SFT           0.12112D-05             0.22615
     X31-NO69          -0.66609D-01            -0.74186
     X57-SO70           0.26956D-03              1.5366
     X60-PA70           0.97647D-04             0.93342
     CONSTANT          -0.51502                 -1.1677

R-SQUARE= 0.5743
SSR=            0.1887D-01      DF= 47
```

Table 4.20

Congenital Birth Defects

```
C\B%   8 -- 61 10 11 21 24 54 25 27 41

     VAR                   B                    T

      X61-DOCH        -0.11009D-02          -0.95740
      X10-NONW        -0.10484              -0.91545
      X11-MAGE         0.90212D-02           1.6360
      X21-DENS         0.51667              0.38505
      X24-COLD         0.41358D-03          0.86881
      X54-CI68         0.11415D-02           1.5569
      X25-XPRO         0.32274D-04          0.60844
      X27-XCAR        -0.15016D-06          -0.26987D-01
      X41-6SFT         0.15405D-04          0.65758
      CONSTANT         -1.3390              -0.61416

R-SQUARE= 0.1867
SSR=           0.5277          DF= 50
```

```
C\B%   8 -- 61 10 11 21 24 54 25 27 41 31 57 60

     VAR                   B                    T

      X61-DOCH        -0.10806D-02          -0.88051
      X10-NONW        -0.62417D-01          -0.50047
      X11-MAGE         0.96362D-02           1.6693
      X21-DENS         0.68174              0.49810
      X24-COLD         0.29908D-03          0.59609
      X54-CI68         0.15099D-02           1.9116
      X25-XPRO         0.36136D-04          0.66141
      X27-XCAR        -0.25284D-05          -0.42856
      X41-6SFT         0.16833D-04          0.60717
      X31-NO69        -0.47663               -1.0255
      X57-SO70        -0.44966D-03          -0.49518
      X60-PA70         0.46820D-03          0.86461
      CONSTANT         -1.3072              -0.57256

R-SQUARE= 0.2205
SSR=           0.5058          DF= 47
```

Table 4.21

Early Infant Diseases

```
  I\B%    9 -- 61 10 11 21 24 54 25 27 41

       VAR                    B                     T

     X61-DOCH           0.39159D-03            0.13339
     X10-NONW           0.89283                 3.0537
     X11-MAGE          -0.44748D-02           -0.31786
     X21-DENS            5.8110                 1.6964
     X24-COLD           0.35758D-03            0.29424
     X54-CI68          -0.13312D-02           -0.71117
     X25-XPRO          -0.10427D-03           -0.76995
     X27-XCAR           0.17641D-04             1.2419
     X41-6SFT           0.15367D-03             2.5695
     CONSTANT           0.68864                0.12373

R-SQUARE= 0.4741
SSR=             3.439          DF= 50

  I\B%    9 -- 61 10 11 21 24 54 25 27 41 31 57 60

       VAR                    B                     T

     X61-DOCH           0.21825D-02            0.73588
     X10-NONW           0.77575                 2.5739
     X11-MAGE          -0.79126D-02           -0.56722
     X21-DENS            5.4464                 1.6467
     X24-COLD          -0.47371D-03           -0.39069
     X54-CI68          -0.12146D-02           -0.63634
     X25-XPRO          -0.13747D-03            -1.0412
     X27-XCAR           0.98932D-05            0.69390
     X41-6SFT           0.67486D-04             1.0073
     X31-NO69          -2.0458                 -1.8215
     X57-SO70           0.43841D-02             1.9978
     X60-PA70           0.17731D-02             1.3549
     CONSTANT            4.1264                0.74792

R-SQUARE= 0.5484
SSR=             2.954          DF= 47
```

Table 4.22

Cancer

```
 CA70  53 -- 61 10 11 21 24 54 25 27 41

      VAR                  B                      T

      X61-DOCH          -0.71271D-02           -3.8356
      X10-NONW           0.63382                3.4249
      X11-MAGE           0.13235                14.854
      X21-DENS            3.9772                1.8342
      X24-COLD           0.17859D-02            2.3216
      X54-CI68           0.50032D-02            4.2228
      X25-XPRO           0.21038D-03            2.4545
      X27-XCAR          -0.13166D-04           -1.4643
      X41-6SFT           0.51683D-04            1.3652
      CONSTANT           -9.5112               -2.6997

R-SQUARE= 0.8556
SSR=           1.378          DF= 50

 CA70  53 -- 61 10 11 21 24 54 25 27 41 31 57 60

      VAR                  B                      T

      X61-DOCH          -0.70763D-02           -3.4976
      X10-NONW           0.64712                3.1475
      X11-MAGE           0.13242                13.915
      X21-DENS            4.0247                1.7837
      X24-COLD           0.17298D-02            2.0913
      X54-CI68           0.51394D-02            3.9471
      X25-XPRO           0.21071D-03            2.3394
      X27-XCAR          -0.14074D-04           -1.4471
      X41-6SFT           0.51447D-04            1.1257
      X31-NO69          -0.16472              -0.21498
      X57-SO70          -0.16143D-03          -0.10784
      X60-PA70           0.23494D-03           0.26317
      CONSTANT           -9.4472               -2.5101

R-SQUARE= 0.8560
SSR=           1.374          DF= 47
```

However, differences between our estimated air pollution effects as opposed to the Lave and Seskin (1977) work are profound. Lave and Seskin (1977) did not find a significant association between particulates and pneumonia. More importantly, Lave and Seskin (1977) found positive associations between air quality (specifically sulfate) and a cardiovascular disease mortality variable and between air quality and cancer mortality. Whether we use SO_2 or the highly collinear sulfate measure, we cannot accept the hypotheses that air pollution has any association with heart and vascular disease or with cancer mortality. Further, our estimated total effects of air pollution on human mortality are about one order of magnitude smaller than those shown by Lave and Seskin (1977).

We can summarize the results of our analysis as follows. When we increase each of the following significant variables by one percent over their mean values in our sample, from the estimated total mortality equation the following percentage change in mean total mortality rate results: (1) for doctors per capita a 0.76 percent decline in mortality rate; (2) for per capita cigarette consumption a 0.32 percent increase in mortality rate; and (3) for per capita protein consumption a 6.7 percent increase in mortality rate. These results suggest several observations. First, medical care, smoking, and diet appear to be enormously important factors in human health. Second, if one looks to a 100% decrease from mean levels for these variables, i.e., the impact on average total mortality of setting these variables to zero, one obtains a 76% increase in mortality for a zero level of doctors per capita, a 32% decrease in mortality for no smoking and a 670% decrease in mortality for no protein in diet. Obviously, the last of these effects is impossible and suggests that we may only have linear approximations of highly non-linear effects. Further, some protein is required to sustain life. Thus, the estimates of mortality effects are likely to be valid only for relatively small changes in explanatory variables. Finally, the air pollution variables are insignificant in the total mortality equation -- as one might suspect if air pollution has only a small effect on mortality rates. This is verified by the fact that the significant estimated effects of particulates on pneumonia and influenze, and of SO_2 on infant diseases are very small in terms of total mortality as compared to the effects of doctors, smoking, and diet.

Given these results, it is important to test the sensitivity of the model to changes in specification of included variables and structure. Two alternative formulations have been specified and tested. First, a version of the model which: (1) uses lagged diet (1955 dietary variable) as opposed to 1965 diet); (2) employs a two-stage doctors per capita variable which includes air pollution in the reduced form equation; and (3) adds lead and sulfate to the air pollution variables, produces essentially identical results both for the impact of medical care and air pollution on mortality. Sulfate air pollution is statistically insignificant across all diseases. The second alternative formulation is identical to the one presented in detail above but the air pollution variables are again included in the reduced form equation for doctors per capita. The results are consistent for the effect of medical care and for the positive associations between sulfur oxides and infant diseases and for particulates and pneumonia. More interesting, however, is a significant negative association which appears between doctors per capita and air pollution in the reduced form equation. It appears that doctors may choose not to live in polluted cities (perhaps for aesthetic reasons). If this is the case, one can easily explain false positive associations between air pollution and mortality where medical care is excluded as an explanatory variable. If doctors avoid polluted cities, and if doctors do reduce mortality rates, then pollution could well be associated with higher mortality rates; but not because of any direct health effect of air pollution on mortality. Rather, failure to account for the locational decisions of doctors (supply and demand for medical care) may well bias estimated epidemiological relationships. In fact, the negative association between doctors per-capita and pollution is so strong, that when pollution is included in the reduced form equation for doctors, the estimated doctors variable used in the two-stage procedure becomes collinear with the pollution variables. This collinearity in some cases produced negative coefficients on the pollution variables in estimated dose-response relationships for some disease categories where pollution is used in the reduced form equation for docotrs per capita. Thus, it is important that, in spite of this collinearity, stable positive associations are retained between pneumonia and influenza and particulates and between infant diseases and sulfur oxides. The inclusion or exclusion of air quality from the reduced form equation has little impact on the conclusions of this study. In part, this occurs because air pollution is collinear with diet. In fact, saturated fats and sulfur oxides are reasonable proxy variables for each other. It has been shown by McCarthy (1971) that the exogenous variables which are collinear with included exogenous variables may be excluded from estimated reduced forms with little loss in consistency in a two-stage least squares procedure.

Another important question for analysis is the possibility that heteroskedasticity is present. At this point, we have only examined one disease category -- cancer mortality -- for this problem. An examination of the residuals plotted against several important explanatory variables (age, for example) showed no evidence of heteroskedasticity.

Finally, in interpreting the results, it should be observed that the associations we have found between mortality and air pollution are principally for diesases of the very young and very old -- particularly susceptible groups within the population. Further, these effects are those which one would perhaps associate with short-term as opposed to long-term air pollution exposures. It may well be that aggregate epidemiology may be incapable of revealing the long-term consequences of air pollution exposures. Two problems are particularly significant here. First, lagged data or data on air pollution histories is not available for such studies. Second, it is nearly impossible to control for population mobility in such studies. Thus, even if one accepts the hypothesis that air pollution levels show enough persistence over time to reveal long-term effects, population mobility will still distort and confound attempts at estimating such effects. A partial remedy for these problems is, of course, to use data on individuals as opposed to aggregate data. The next chapter provides a preliminary exploration of just such a data set.

We now turn to an economic evaluation of the value of air pollution control in reducing mortality based on the value of safety approach described

Table 4.23

Methodology for Health Benefits Assessment

Benefits = (Population at Risk) x (Value of Safety) x (Reduction in Health Risk)

Value of Safety Based on Consumer's Willingness to Pay

Low estimate: $340,000

Source: Thaler & Rosen (1975)

High Estimate: $1,000,000

Source: Robert Smith (1974)

above.

4.6 A Tentative Estimate of The Value of Safety from Air Pollution Control

Given all of the caveats discussed above concerning the validity of the estimated effects of air pollution on mortality, it is possible to construct benefit measures using the methodology outlined in Section 4.2 above. The methodology is briefly summarized in Table 4.23.

First, to obtain national estimates, we must know the population at risk. Since our sixty-city sample is entirely urban, and since air pollution is principally an urban problem we will use a population risk for 1970 of 150 million urban dwellers. As a range for the value of safety, we will employ Thaler and Rosen's (1975) estimate of $340,000 (in 1978 dollars) as a lower bound and Smith's (1974) estimate of $1,000,000 (in 1978 dollars) as an upper bound. Finally, to provide an estimate of reduced risk from air pollution control, we will assume an average 60% reduction in ambient urban concentrations both for SO_2 and particulates. Then, using the mean concentration of these pollutants in our sixty-city sample as a basis for calculation, we can derive the average reduction in risk of pneumonia mortality for a 60% reduction in particulates and the average reduction in risk of infant diseases for a 60% reduction in SO_2 from our estimated dose response functions for these diseases.

Multiplying the population at risk by the assumed value of safety, and then by the average reduction in risk, gives a crude approximation of the benefits for a 60% reduction in national urban ambient concentrations of particulates and SO_2, respectively. National urban totals and the value of the average individual risk reduction are shown in Table 4.24.

The value estimates as shown in Table 4.24 agree surprisingly well with those developed by Lave and Seskin (1977) for national air pollution damages. However, the dollar value is similar only because we use a range for the value of safety (derived from observed market behavior of consumers) which is about an order of magnitude larger than the "value of life" figure based on lost earnings which is employed by Lave and Seskin (1977). We, of course, reject the value of life notion, instead focusing on the measurable concept of value of safety. Since there is no evidence to suggest that society puts less value on safety for children, the aged or women than on employed heads of households, we feel that the best measures available now for the value of safety should be employed for all individuals. Eventually, more refined measures of the value that different individuals place on safety may become available. However, for the time being, these are the best valuations of the social worth of safety we can employ.

Table 4.24

Urban Benefits from Reduced Mortality: Value of Safety for 60% Air Pollution Control

Disease	Pollutant	Average Individual Safety Benefit (1978 Dollars/Year)	National Urban Benefits (1978 Billion Dollars/Year)
Pneumonia	Particulates	29 - 92	4.4 - 13.7
Early Infant Disease	SO_2	5 - 14	.7 - 2.2
Total		34 - 106	5.1 - 15.9

CHAPTER V

THE MICHIGAN SURVEY EXPERIMENT

5.1 Objectives of the Experiment

The data set employed in this chapter refers to the health status and the time and budget allocations of each of several thousand household heads over a nine-year period. Its highly disaggregated form therefore avoids many of the estimation problems associated with the aggregate data used in Chapter IV. This avoidance is not our only purpose, however. The richness of detail in the data set allows us to extend the range of phenomena that we study. Most important, we are able to investigate the morbidity effects of air pollution, considering acute effects and chronic effects separately. The detail of the data set allows us to identify much more readily those variables that are not current determinants of health status, thus providing a means of avoiding the simultaneity problems that plagued the aggregate dose-response functions of the previous chapter. It is important to note that the results reported here reflect a preliminary attempt to evaluate the usefulness of Michigan Survey Data in estimating morbidity (sickness) effects of air pollution and consequent economic losses. As a result of the preliminary nature of the research, many highly desirable transformations of the variables as defined in the Michigan Survey Data set have not yet been made. However, in spite of the preliminary nature of the results they do represent the first attempt to qualify the economic losses due to morbidity as opposed to mortality resulting from air pollution.

With the richness of the data available to us, we need not terminate our efforts after having estimated a set of dose-response expressions for the morbidity effects of air pollution. We are able to ascertain the labor productivity effects and the impact on willingness to pay to avoid chronic and/or acute illness as well. Both of these additional efforts are undertaken in this chapter.

5.2 The Sample and the Variables

Our analysis is based on yearly interviews conducted by the University of Michigan's Survey Research Center with a nationwide random sample of 4,802 to 5,862 families from 1968 through 1976. No families with living members were ever intentionally deleted from the sample, and, as families broke apart, the adult components were added to the sample as distinct families. The cumulative interview response rate over the nine-year period declined from 76 percent in the 1968 and first interview year to 55 percent in the 1976 interview year, implying an average yearly reinterview response rate of nearly 95 percent. From 1970 through 1976, this yearly response rate averaged 97 percent. Of special interest to us is that, in addition to substantial detail on household head time and budget allocations, the sample contains generalized measures of the head's health states as well as information on lifestyle and biological and social endowment variables that might plausibly contribute to the health states.

Information from the interview has been combined with data on a limited set of environmental variables, particularly information on air pollution concentrations, to establish imperfect measures of the environment in which each family head has lived during the nine-year period. To the best of our knowledge, the Survey Research Center data set is the only one currently available that combines, for the same set of individuals over a substantial number of years, information on places of residence, states-of-health, and time and budget allocations. The sample thus raises the prospect of our being able to value, through empirical applications of the economic theory of consumer behavior, the contributions of environmental pollution exposures to states-of-health.

The major characteristics of our sample and the variables we employ in our empirical efforts are presented in Tables 5.1 and 5.2. All variables refer to household heads. Table 5.1 gives complete definitions of variables, their scalings, and their assigned acronyms; Table 5.2 provides representative arithmetic means and standard deviations of variables used. Because we employ various partitions of the sample throughout the chapter, we do not use the Survey Research Center sample weights. Our samples are therefore not entirely representative of the national population.

In Table 5.2a, so as not to make worse the already considerable and cumbersome length of the listing, only the two health variables, LDSA and ACUT, are listed as dependent variables. The geometric means of the air pollution variables have their means and standard deviations entered for the various sample partitionings indicated at the bottom of the table. The means and standard deviations for the other variables are listed in Table 5.2b. This latter table refers only to the samples used for the chronic illness expressions, while the former refers to the acute illness expressions. Whether reference is to the partitioned or unpartitioned samples, the means and standard deviations represent only those samples used to estimate dose-response functions involving geometric mean measures of the air pollution variables. All estimates employing different combinations of variables, whatever the combination might be, were established using a random drawing from the entire Survey Research Center population sample for a particular year. Therefore, the means and standard deviations listed in Table 5.2, although extremely representative, are not the exact values for each of the samples used in the estimation effort.

The definition and measurement of most of the variables listed in Tables 5.1 and 5.2 is standard, and we shall comment here only on those that pose definitional and measurement problems for the major focus of this report. This criterion immediately directs attention to the air pollution variables.

Table 5.1

Complete Variable Definitions

Health State Variables

Acute illness (ACUT) -- workdays ill times 16 for the first 8 weeks and times 12 thereafter. Only individuals who are currently employed or unemployed and looking for work could have positive values for this variable.

Degree of disability (DSAB) -- complete limitation on work = 1; severe limitation on work = 2; some limitation on work = 3; otherwise = 0.

Length of disability (LDSA) -- $\leq$ 2 years = 1; 2 - 4 years = 2; 5 - 7 years = 3; $\geq$ 8 years = 4; otherwise = 0. This is a follow-up question to inquiries about whether the respondent has any physical or nervous condition that limits the amount or kind of work or housework he can do.

Biological and Social Endowment Variables

Age of family head in years (AGEH)

Grew up in city (CITY) = 1; otherwise = 0. This variable, as transformed, is binary.

Education attainment (EDUC) -- 6 - 8 grades = 2; 9 - 11 grades = 3; 12 grades = 4; 12 grades plus non-academic training = 5; college, no degree = 6; college degree = 7; advanced or professional degree = 8; otherwise = 1.

Father's educational attainment (FEDU) -- same scaling as for EDUC.

Family size in number of persons in housing unit (FMSZ).

Length of present employment (LOCC) -- $\leq$ 1 year = 1; 12 - 19 months = 2; 1-1/2 - 3-1/2 years = 3; 3-1/2 - 9-1/2 years = 4; 9-1/2 - 19-1/2 years = 5; $\geq$ 19-1/2 years = 6; otherwise = 0.

Marital status (MARR) -- married = 1; otherwise = 0. This variable, as transformed, is binary.

Income level of parents (POOR) -- poor = 1; otherwise = 0. This question asked whether the respondent's parents were " . . . poor when you were growing up, pretty well off, or what?" The variable, as transformed, is binary.

Race of family head (RACE) -- white = 1; otherwise = 0. This variable, as transformed, is binary.

Sex of family head (SEXH) -- male = 1; otherwise = 0. This variable, as transformed, is binary.

Member of a labor union (UION) -- Yes = 1; otherwise = 0. This variable, as transformed, is binary.

Life Style Variables

Practices absenteeism from work (ABSN) -- absent once or more a week from work = 1; otherwise = 0. This refers to a question in which the respondent is asked if there are times when he doesn't go to work at all, even if he isn't sick. The variable, as transformed, is binary.

Table 5.1
(continued)

Frequency of church attendance (CHCH) -- once a week or more = 1; otherwise = 0. This variable, as transformed, is binary.

Annual family expenditures on cigarettes in dollars (CIGE) -- this variable is *not* indexed for differences in prices among locales.

Participates in energetic activities (EXER) -- first mention = 1; otherwise = 0. This question asks the family head what he usually does in his spare time. Energetic activities include fishing, bowling, tennis, camping, travel, hunting, dancing, motorcycling, etc.

Family food consumption relative to food needs standard in percent (FOOD) -- family food consumption refers to food expenditures in dollars and includes amounts spent in the home, school, work, and restaurants, as well as the amount saved in dollars by eating at work or school, raising, canning or freezing food, using food stamps, and receiving free food. The food needs standard is in dollars and is based on USDA Low Cost Plan estimates of weekly food costs as published in the March 1967 issue of the *Family Economics Review*. The standard itself is calculated by multiplying the aforementioned weekly food needs by 52 and making a series of adjustments according to the size of the family.

Is often late to work (LTWK) -- late once or more a week to work = 1; otherwise = 0. This question asks the respondent if there are times when he is late getting to work. The variable, as transformed, is binary.

Daily number of cigarettes smoked per adult family member (NCIG) -- $\leq$ 3 = 1; 3 - 17 = 2; 18 - 22 = 3; 23 - 35 = 4; 2 - 3 packs = 5; $\geq$ 4 packs = 6; otherwise = 0.

Fundamentalist religious preference (RELG) -- Mormon, United Church of Christ, Disciples of Christ, Quaker, etc. = 1; otherwise = 0. This variable, as transformed, is binary.

Degree of risk aversion (RISK) -- a weighted index devised by the survey team in which the individual's degree of risk aversion increases if he drives the newest car in good condition, does not own a car, has all cars insured, uses seat belts, has medical insurance, smokes less than a pack a day, has some liquid savings, and has more than two month's income saved. Nine is the greatest degree of risk aversion that can be exhibited.

Head's annual hours working for money (WORK).

Precuniary Variables

Cost-of-living in 1970 country of residence (BDALO) -- an index of comparative costs for a four-person family living in various areas as published by the U.S. Bureau of Labor Statistics in the Spring 1967 issue of *Three Standards of Living for an Urban Family of Four Persons*. The lowest living standard was employed. This index is published for the thirty-nine largest SMSA's and by region for the nonmetropolitan areas. For the remaining SMSA's, the regional average of the metropolitan indices was used.

Table 5.1
(continued)

Has hospital or medical insurance (INSR) -- Yes = 1; otherwise = 0. This variable, as transformed, is binary.

Family income in dollars not due to current work effort (ICTR) -- this variable includes assorted welfare payments, pensions, and annuities, as well as earnings from assets.

Family net real income in dollars (RINC) -- this variable is the sum of money income plus value of goods and services received at less than market prices less the cost of earning income.

Savings in dollars equal or greater than two month's income (SVGS) -- Yes = 1; otherwise = 0.

Head's marginal hourly earnings rate in cents (WAGE) -- in circumstances where the head neither has a second job nor commands overtime pay, this variable is simply total annual earnings from labor divided by annual hours worked for money. Where he has two or more jobs, it is his hourly earnings in the last job he names. If he has only one job, can and does work overtime if he wishes, and receives overtime pay, the variable is his average overtime hourly earnings.

Environmental Variables

Works in chemicals or metals manufacturing industries (CHEM) -- Yes = 1; otherwise = 0. The chemicals industry includes chemicals and allied products, petroleum and coal products, and rubber and miscellaneous plastic products. The metals industry includes steel, aluminum, foundaries, etc.

Number of days in 1972 when temperatures were below freezing at some time during the day (COLD). This data was obtained from USNOAA, Climatological Data, National Summary 1972.

Number of persons per room in family dwelling (DENS).

Distance from nearest city of 50,000 or more people (MILE) -- $\leq$ 5 miles or outside continental United States = 1; 5 - 15 miles = 2; 15 - 30 miles = 3; 30 - 50 miles = 4; $\geq$ 50 miles = 5.

Nitrogen dioxide: annual 24-hour geometric mean (M), ninetieth percentile (N), and 30th percentile (T) in micrograms per cubic meter as measured by the Gas Bubbler TGS Method-Frit before 1974 and the Saltzman method for 1974 and after (NOX). This data was obtained from the annual USEPA publication, Air Quality Data -- Annual Statistics.

Sulfur dioxide: annual 24-hour geometric mean (M), 90th percentile (N), and 30th percentile (T) in micrograms per cubic meter as measured by the Gas Bubbler Pararosaniline-Sulfanic Acid Method (SUL). This data was obtained from the annual USEPA publication, Air Quality Data - Annual Statistics.

Total suspended particulates: annual 24-hour geometric mean (M), 90th percentile (N), and 30th percentile (T), in micrograms per cubic meter as measured by the Hi-Vol Gravimetric method (TSP). This data was obtained from the annual USEPA publication, Air Quality Data -- Annual Statistics.

Table 5.1
(continued)

Ultraviolet radiation in microwatts per square centimeter (ULTV). This data was taken from Pazand, R., Environmental Carcinogenesis -- An Economic Analysis of Risk, unpublished Ph.D. dissertation. University of New Mexico (June 1976).

Explanation of Table

Unless otherwise stated, all data is taken from tapes described in Survey Research Center, A Panel Study of Income Dynamics, Ann Arbor: Institute for Social Research, University of Michigan (1972, 1973, 1974 1975, 1976).

All variables referring to an individual person refer only to the family head.

On occasion, definitions for the same phenomenon will differ from year to year. If this occurs, a single integer indicating the year to which reference is made is attached to the end of the variable acronym. Thus 1967 = 7; 1968 = 8; . . .; 1976 = 6.

Table 5.2a

Representative Means and Standard Deviations of Health and Air Pollution Variables for Samples Involving Family Heads Currently Employed or Actively Looking for Work*

Variable Acronym	Year							
	1967	1968	1969	1970	1971	1972	1973	1974[b]
Health States								
ACUT	100.414	120.486	133.657	113.750	113.323	149.845	112.530	
	(183.594)	(214.759)	(332.171)	(277.022)	(266.274)	(427.983)	(259.120)	
LDSA[a]	0.953	0.645	0.337	0.363	0.268	0.290	0.260	0.348
	(1.720)	(1.326)	(0.979)	(0.971)	(0.888)	(0.921)	(0.874)	(0.952)
Environmental								
NCXM						157.043	118.045	
						(51.070)	(72.230)	
SULM	24.475	25.113	27.220	16.286	17.657		2.051	7.435
	(19.098)	(18.714)	(25.013)	(12.150)	(9.449)		(4.188)	(11.728)
TSPM	100.403	99.917	98.713	95.534	87.213	99.157	35.310	71.108
	(35.469)	(30.628)	(29.609)	(18.943)	(27.920)	(30.941)	(42.183)	(36.085)

[a] Except for 1970, all samples refer to family heads who have never lived in more than than one state. In 1971, the reference is to family heads who currently live within walking distance of relatives.

[b] Includes housewives, retirees, and students.

*Standard deviations are enclosed in parentheses.

Table 5.2b
Representative Means and Standard Deviations of All Other Variables[a]

Variable Acronym	1967	1968	1969	1970	1971	1972	1973	1974	1975
Health State									
DSAB	0.493 (1.291)		0.111 (0.315)	0.426 (1.634)	0.488 (1.011)	0.470 (0.949)	0.304 (0.754)	0.800 (2.159)	0.624 (1.854)
Biological and Social Endowment									
AGEH	43.558 (12.337)		40.323 (11.841)	43.745 (13.451)	44.218 (13.649)	44.305 (15.276)	45.155 (16.158)	37.322 (15.421)	37.925 14.749
CITY	0.646 (0.481)		0.451 (0.498)	0.678 (0.468)	0.678 (0.468)	0.632 (0.459)	0.655 (0.476)	- -	- -
EDUC	3.680 (1.696)		3.683 (1.747)	3.878 (1.862)	3.923 (1.866)	7.705 (1.851)	3.720 (1.844)	3.912 (1.672)	3.659 (1.685)
FEDU	2.391 (2.254)		2.300 (2.036)	2.313 (1.442)	2.360 (1.473)	2.458 (1.609)	2.395 (1.451)	- -	- -
FMSZ	3.812 (2.401)		4.586 (2.542)	3.993 (2.376)	3.930 (2.412)	3.508 (2.202)	3.233 (2.126)	- -	- -
LOCC	2.257 (2.234)		3.271 (1.869)	- -	- -	2.283 (2.168)	2.168 (2.188)	- -	- -
MARR	0.617 (0.489)		0.617 (0.487)	- -	- -	- -	0.525 (0.500)	0.468 (0.500)	0.540 (0.500)
POOR	0.578 (0.496)		0.543 (0.499)	- -	0.520 0.500	0.490 (0.501)	0.520 (0.500)	0.551 (0.499)	0.615 (0.488)

(continued)

Table 5.2b
(continued)

Variable Acronym	1967	1968	1969	1970	1971	1972	1973	1974	1975
RACE	0.469 (0.501)		0.917 (0.276)	0.410 (0.099)	0.500 (0.501)	0.443 (0.497)	0.475 (0.500)	- -	0.346 (0.477)
SEXH	0.629 (0.468)		0.677 (0.496)	0.635 (0.482)	0.635 (0.482)	0.573 (0.495)	0.603 (0.490)	0.640 (0.382)	0.631 (0.417)
UION	- -		0.354 (0.479)	0.233 (0.423)	0.198 (0.399)	- -	0.198 (0.399)	- -	- -
Lifestyle									
ABSN	- -		- -	- -	0.108 (0.310)	- -	- -	- -	- -
CHCH	- -		0.440 (0.448)	- -	- -	- -	- -	- -	- -
CIGE	-		93.146 (124.022)	- -	- -	- -	- -	- -	- -
EXER	0.144 (0.352)		0.189 (0.392)	0.225 (0.418)	0.198 (0.399)	- -	- -	- -	- -
FOOD	505.830 (380.977)		757.669 (372.594)	822.500 (716.450)	840.990 (716.100)	- -	- -	1030.976 (574.163)	1145.150 (707.099)
LTWK	- -		- -	0.070 (0.255)	0.209 (0.407)	- -	- -	- -	- -

Table 5.2b
(continued)

Variable Acronym	1967	1968	1969	1970	1971	1972	1973	1974	1975
NCIG	1.851 (1.912)		- -	- -	- -	- -	- -	- -	- -
RELG	0.018 (0.136)		- -	- -	- -	- -	- -	0.054 (0.226)	0.062 (0.242)
RISK	4.489 (1.605)		4.503 (1.452)	4.658 (1.545)	4.673 (1.540)	- -	- -	- -	- -
WORK	1245.875 (1059.780)		1989.649 (674.723)	1560.895 (1001.253)	1527.732 (982.381)	1333.540 (1030.346)	1354.137 (1056.153)	- -	- -
Pecuniary									
BDALO	99.638 (4.720)		99.220 (4.297)	100.413 (4.625)	100.266 (4.788)	100.618 (4.925)	100.736 (4.819)	- -	- -
INSR	0.889 (0.316)		0.794 (0.404)	0.708 (0.455)	0.695 (0.461)	- -	- -	- -	- -
ICTR	1096.22 (1314.401)		508.249 (1124.259)	1238.392 (1198.698)	1013.846 (1721.377)	1342.585 (1874.235)	1366.702 (1993.720)	-	-
RINC	9148.605 6511.900		8902.377 (6100.167)	10852.230 (7833.473)	10875.650 (7439.632)	9556.803 (7274.871)	11077.950 (8337.711)	- -	- -

(continued)

Table 5.2b
(continued)

Variable Acronym	1967	1968	1969	1970	1971	1972	1973	1974	1975
SVGS	0.342 (0.475)		0.289 (0.454)	0.333 (0.472)	0.371 (0.484)	- -	- -	- -	- -
WAGE	292.119 (405.985)		314.440 (221.346)	322.500 (316.450)	358.258 (331.738)	298.230 (319.890)	336.525 (337.425)	- -	- -
Environmental									
CHEM	0.022 (0.147)		- -	0.008 (0.086)	- -	- -	0.003 (0.050)	0.049 (0.216)	0.045 (0.206)
COLD	81.502 (52.684)		- -	- -	- -	- -	- -	- -	- -
DENS	- -		3.420 (1.797)	- -	- -	0.870 (1.198)	0.725 (0.414)	- -	- -
NOXN	- -		- -	- -	- -	246.573 79.826	104.860 (75.994)	97.429 (44.564)	90.717 (22.716)
NOXT	- -		- -	- -	- -	132.045 (37.087)	31.536 (23.964)	32.931 (31.761)	48.597 (13.911)
SULN	107.687 (134.484)		- -	74.663 66.016	61.768 (38.495)	- -	42.625 (31.115)	34.566 (42.841)	25.650 (41.603)

(continued)

Table 5.2b
(continued)

Variable Acronym	1967	1968	1969	1970	1971	1972	1973	1974	1975
SULT	26.041 (37.369)			10.798 (10.663)	11.190 (5.875)	- -	9.551 (9.305)	5.006 (9.955)	7.836 8.233
TSPN	176.986 (78.097)			248.965 (339.668)	156.185 (63.787)	170.768 (58.121)	147.960 (39.684)	126.702 (43.086)	120.580 (56.438)
TSPT	77.605 (23.661)			74.837 (43.932)	74.088 (20.772)	82.995 (26.627)	56.232 (9.650)	67.122 (22.200)	62.779 (27.046)
ULTV	1494.75 (634.638)			- -	- -	- -	- -	- -	- -

[a] All samples include housewives, retirees, and students.

*Standard deviations are in parentheses.

If one has detailed and real-time information on changes in health states, ideally one would like to have real-time records of all air pollution exposures. The coarse yearly indicators of acute and chronic illness in the Survey Research Center (SRC henceforth) data could not support such detail. We therefore chose to collect outdoor air pollution data averaged over a time period corresponding to the time interval employed in the SRC data. In addition, we wished to ascertain whether representations of moments of the outdoor air pollution frequency distribution other than measures of central tendency might contribute to ill-health. The result of these deliberations was a decision to acquire data on the geometric mean (because outdoor air pollution tends to be log-normally distributed over time), 30th percentile, and 90th percentile of the annual concentrations of five pollutants: nitrogen dioxide; ozone; total oxidants; total suspended particulates; and sulfur dioxide. Although the ozone and total oxidant data has been combined with the SRC data, the number of monitoring locations and the monitoring time intervals were inadequate to allow other than minor variations in the exposures of the sample individuals. Thus the empirical results to be reported neglect these two possibly important pollutants.

Matching the thousands of outdoor air pollution monitoring stations in the United States to the hundreds of counties where the SRC sample families resided could be a complex combinatorial problem. The matching was achieved for each of the nine years at the cost of not having outdoor air pollution information for some SRC sample families during some years and of assigning somewhat inappropriate air pollution exposures to some sample individuals. The full extent of this information loss is presently unknown.

The matching process started by listing all the counties in the United States where one or more SRC sample families had resided during the nine year interval. Separately for each of the five previously mentioned air pollutants, a yearly listing of the counties having outdoor air pollution data for one or more of the three frequency distribution measures being considered was constructed. Of the 301 counties in 50 states where sample families resided during the nine year interval, outdoor air pollution monitoring data for one or more of the measures of one or more of the five air pollutants existed at least for one year in 118 of the counties in 50 states. No attempts were made to extrapolate air pollution data from one county to another, nor were any switches between monitoring stations in a single county ever made. In counties where multiple outdoor monitoring stations were present, the data from the single station that had operated for the greatest portion of the nine years was used. If two or more stations in a county had operated for equal portions of the nine years, the station having the most complete (in terms of numbers of pollutants and pollutant measures) was employed. When air pollution data were available in a family's residence county for a particular year, these criteria served to assign outdoor air pollution exposures to all sample families. For most years, somewhat more than 3,000 families had some sort of outdoor air pollution data assigned them. Because of our reluctance to adopt a new monitoring station location in a county whenever the activities of a station we had previously used were terminated, we undoubtedly missed a few opportunities to assign air pollution data to a few sample families. This issue pales, however, beside the issue of the extent to which the assigned data represent actual outdoor air pollution exposures.

The SRC family data sample provides only the family's county and state of residence: it does not give the home town or city. Thus, for large urban counties such as Cook County, Illinois, or Los Angeles County, California, or occasional rural counties such as San Bernardino County, California, where there exist major locational differences in potential air pollution exposures within the county, substantial error could exist in the air pollution assignations. This important source of measurement error could perhaps be substantially reduced if all counties having this property were identified and if all families residing in the identified counties were excised from the sample. We have made no attempt to perform this excision.[1/]

This criteria employed to select pollution monitoring stations probably result in the assignment of downtown urban locations, where pollution concentrations have historically tended to be greatest and where the most extensive monitoring has been done. Since relatively few of the SRC sample families actually live in downtown areas, the constructed data set generally exaggerates family outdoor air pollution exposures, implying that the health effects, if any, of air pollution will tend to be underestimated.[2/]

Outdoor air pollution at the place of residence is not the only plausible environmental source of deleterious health effects. Indoor air pollution at home and in the work place, outdoor air pollution at other locations, contaminants in diet, and water pollution are additional widely acknowledged possible sources. We introduce measures (albeit imperfect) of some of these plausible alternative sources in our empirical efforts and fail to give any attention to others such as water pollution. If these excluded types of pollution have health effects of their own, and if their extent tends to be positively correlated with the extent of outdoor air pollution, then the included air pollution variables will capture some of their contributions to ill-health, causing the measured contribution of the outdoor air pollution variables to be exaggerated. The extent of this upward bias will vary directly with the degree of correlation between the included and the excluded variables and the extent to which the excluded variable actually contributes to the effect of interest. For this study, of the previously mentioned alternative environmental pollution sources of health effects, the utter exclusion of any measures of water pollution is perhaps the most serious. At various points in the empirical effort, rather crude measures of indoor home air pollution (family smoking habits), diet (a dietary adequacy index), and indoor air pollution at the work place (employment in the chemicals or metals manufacturing sector) are included.[3/]

The issue of excluding possibly relevant variables from the analysis included outdoor air pollution as well. Oxidants and ozone, because of insufficient variation in apparent exposures among sample families, have been disregarded, even though exposure values are present in the constructed data set. Other important air pollutants, for which data were available such as carbon monoxide, were not even considered because of the large variations in their instantaneous concentrations within a few hundreds of feet. Some pollutants that have attracted recent regulatory and public concern, such as acid sulfates, had no data readily available. Finally, of the pollutants that were included in the constructed data set and exploited in the empirical effort, the time series for all except total suspended particulates were incomplete. Thus, for example, no information was available on sulfur dioxide concentrations in 1972.

Measurement error is not only an issue in the outdoor air pollution variables. What some might choose to interpret as measurement error is a prime feature of the two dependent variables, number of days annually ill and length of time disabled.4/ Although we have no basis other than seemingly sensible intuitive interpretations of the form of the questions asked the respondents (see the explanations for ACUT and LDSA in Table 5.1), we choose to interpret the former as acute illness and the latter as chronic illness. Definitional problems of the distinction between acute and chronic illness aside, it must be remembered that what is an illness to one individual is not an illness to another individual. Even the same individual may differ over time in what he considers to be a state of illness. Illness is, in part, an idiosyncratic and subjective phenomenon only partly susceptible to consensus standards of definition. Therefore, if one prefers a reductionist perspective and wishes to have all phenomena collapse to, say, a chemical measurement, then the values of the variables we are trying to explain in this study indeed leave a great deal to be desired.5/ Economic analysis, however, presumes that illness and its costs lie in the eye of the beholder. No laws whatsoever governing choices are innate in the material objects of ordinary cognition. As has been emphasized in the introduction to this section of the report, the degree of illness that afflicts an individual is, in part, often a matter of purposive choice. Economic principles relate to the subjective desires motivating individuals to become aware of and perhaps to alter their environments. Thus no object or status becomes relevant in economic analysis until humans perceive it can be used for or defeats some subjective purpose. Illness that is defined in clinical terms but which is never subjectively realized by the individual who is said to be clinically ill is of little interest except to clinicians. It is certainly arguable whether their standards of what constitutes illness should prevail over those of the individual who professes illness. For this study, we are forced by circumstances to adopt the latter's perspective. Fortunately, it fits readily into economic analysis.

In spite of the preceding argument a type of measurement error does persist in the two dependent health variables. This type of error is inherent in the use of any fairly encompassing measure of health status. Kinds of debilitating acute illness for an individual may range, for example, from headaches to heel blisters. Chronic illnesses may show similar variations over body sites and implied debilitating effects for the same individuals. In effect, therefore, an individual's response to a question about the number of days he has been ill or the length of time he has been disabled involves an aggregation of several attributes perhaps sampled from some larger population of attributes. The weights the respondent employs to combine these attributes to obtain the encompassing health measures may differ among individuals. Furthermore, they may not be those weights that correspond to the contribution of the attribute to some other parameter of interest, such as hours of work or money wages. Recognition of the possibility that individuals may employ different weights to aggregate to the encompassing health measure serves perhaps to deepen the reader's perception of the subjectivity of our measures of ill-health. It says only that there may be as many unique measures of ill-health employed as there are respondents in the sample. The import for our empirical efforts of discrepancies between the contributions of attributes to ill-health and to other parameters of interest is greater, since we shall try to ascertain the impact of air pollution-induced ill-health upon labor supply and productivity. In particular, the use of the encompassing measures of ill-health rather than the specific attributes may attenuate our estimate of the effect of air pollution-induced health effects upon labor supply and productivity.

As Table 5.1 indicates, all SRC sample individuals not currently employed or seriously looking for current employment had no information recorded about the number of days they professed to be acutely ill. Furthermore, those individuals for whom information on ACUT was recorded were never sick on weekends: their accute illneses occurred, according to the data, only on workdays. The ACUT variable may thus be confounded by the wish of some respondents to legitimatize for the sake of social appearance or internal self-respect their failure to go to work. In the empirical efforts regarding ACUT therefore, an actual choice of leisure over labor could thus be falsely attributed to ill health. Marquis (1978), however has been unable to discover any basis for this source of bias.

The rather long list of other variables considered can be divided, somewhat imperfectly, into health state, biological and social endowment, lifestyle, pecuniary, and environmental variables. For the moment, we will limit our discussion of the variables not already discussed to the parts they are expected to play in dose-response functions, reserving the discussion of labor supply and productivity impacts to a later section. Only those variables actually used in the estimated dose-response functions are therefore discussed in this section. A summary table of expected signs is presented in Table 5.3.

DSAB, the degree of disability is the only included health state variable not employed as a dependent variable. Since it is ordinally scaled, its meaning when used as a dependent variable is arbitary. Any four or five monotonically increasing numbers would have no more and no less meaning. When entered as an explanatory variable in the chronic illness production function, its expected sign is unclear. If the individual continues to live in spite of having a chronic disability, one would expect the period of recovery, if any, to be lengthier the more severe the disability. However, in the general population, severe disabilities perhaps are more likely to lead to earlier death. Thus, those sample individuals who are severely disabled might be expected to have been disabled only for a relatively short time span. This would lead one to expect a negative association between DSAB and LDSA. Which effect would dominate in any particular sample must be conjectural. In contrast, since disabilities, both in terms of length and severity, probably cause the

Table 5.3

Expected Signs for Explanatory Variables in Estimated Dose-Response Functions

	Acute Illness	Chronic Illness
Health States		
DSAB	+	?
LDSA	+	X
Biological and Social Endowments		
AGEH	+	+
CITY	?	?
EDUC	?	?
FEDU	−	−
FMSZ	−	?
MARR	−	?
POOR	+	+
RACE	−	−
SEXH	?	?
Lifestyles		
CHCH	−	−
EXER	−	X
FOOD	−	−
NCIG	+	+
RELG	−	−
RISK	−	−
Precuniary		
INSR	−	−
Environmental		
CHEM	+	+
COLD	?	?
DENS	+	+
All NOX	+	+
All SUL	+	+
All TSP	+	+
ULTV	?	?

? ≡ unknown
X ≡ irrelevant

individual to be more susceptible to common temporary illness, we expect LDSA and DSAB to contribute positively to ACUT. However, because the values for DSAB are not monotonically ordered, the magnitudes of the coefficients for DSAB in both the LDSA or the ACUT expressions should be disregarded.

No one holds that health states improve with adult age. The adult human organism suffers natural decay, making the investment necessary to maintain a given health state progressively more costly. The inclusion of two additional irrevocable biological attributes, race and sex, can be justified on at least two grounds. First, susceptibilities to some diseases differ by race or sex. Men, for example, don't have breast surgery and whites don't contact sickle cell anemia. The implications of this for the signs of RACE and SEXH are unclear, however. Second, and probably most important with respect to race, minorities have frequently had less preventive and ameliorative medical care available to them and have perhaps had less opportunity to learn how to use what is available wisely. The RACE variable might therefore capture some fair portion of past and present differences in the availability of medical services to individuals. If this speculation is correct, RACE, which has a value of 1 if the individual is white and 0 otherwise, should have a negative sign attached for both illness types.

CITY, FEDU, and POOR are intended to represent differences among individuals in their childhood environments. If one grew up in a city, he probably had better access to medical care. On the other hand, he was probably exposed to more toxics in his everyday environment. The sign to be expected for CITY is therefore ambiguous. In contrast, the proper signs to expect for FEDU and POOR are relatively unambiguous. Educated parents, in addition to their other knowledge about worldly affairs, will perhaps be more sensitive to the implications of childhood health practices for future adult health status of the child. In addition, they might tend to be better at interpreting signals of health distress and choosing the medically most effective course of action. If adult health states are positively influenced by childhood health practices, then the sign attached to the FEDU coefficients in either acute or chronic illness dose-response functions should be negative. For similar reasons, the POOR coefficients are expected to have positive signs.

With one ambiguous exception, EDUC, FMSZ, and MARR contribute to good health. Many recent studies indicate that among socioeconomic variables, years of formal schooling completed is frequently the most important predictor of good health. Grossman (1975) has found empirical evidence of a causal relationship running from past schooling to current health. The individual who is married has his wife's time available, as well as his own, for the protection of his health. At least for acute illness, increasing family size also implies that certain individuals within the family can specialize in the production and the protection of other family members' health. This implies that over some interval there exist increasing returns to health production specialization within the family, a proposition that accords neatly with casual observation but for which no strong empirical evidence appears to exist.

The expected sign for FMSZ in a chronic illness dose-response function is ambiguous because the number of children a family has is, in part, an investment decision.[6] Older children provide more opportunities for family members to specialize in health production and protection; however, if a state of chronic disability was suffered by the family head before the accumulation of a large family, it would seem that the investment process in children would be made more costly. The latter statement implies that fewer children and chronic disability are positively associated, while the former says that children, once they are able to assume some responsibilities for family production, contribute to good health. Put in terms of our concerns in the introduction to this portion of the report, an observed association between an individual's state-of-health and his family size could reflect causality running both from family size to health and from health to family size. This issue could, of course, be resolved by building an analytical structure in which family size is made a decision variable. To do so would take us beyond the immediate scope of this research effort. We have therefore employed family size as an explanatory variable in our estimated chronic illness dose-response functions without imposing any sign expectations upon it and recognizing that its presence could bias the air pollution coefficients.

All of the lifestyle explanatory variables are standard entries in epidemiological studies of air pollution. There are, however some special features worthy of note for each variable. NCIG, for example, is not the number of cigarettes smoked by the individuals but rather the number smoked per adult family member. It is assumed this serves as a reasonable proxy for the smoking habits of the individual head. For the cigarette variable therefore, its estimated coefficient is best considered as an indicator of the health effects of smoking or not smoking. Little, if any, credence should be assigned these coefficients as indicators, in the neighborhood of the average smoking habits of the respondent sample, of the incremental health effects of smoking an additional cigarette; that is, the sign of the coefficient rather than its magnitude is the result to inspect.

Biomedical wisdom says that continuing participation in energetic activities and an adequate diet contribute to good health. Since the SRC data set contained no information on the respondent's exercise habits before he became disabled, we have not included EXER in the chronic illness dose-response function. Otherwise one must face the two-way causality problem with inadequate data resources to handle it. In neglecting this variable, however, which proves to be consistently statistically significant in the acute illness dose-response function, we raise the spectre of biasing the air pollution coefficients in the estimated chronic illness dose-response functions. Since, a priori, energetic activities are expected to reduce the incidence of chronic illness, the absolute magnitudes of the air pollution coefficients will be biased downward, causing the effect of air pollution on chronic illness to be underestimated. However, for those years in which EXER is available in the SRC data set, the absolute value of the simple correlation between it and the air pollution variables is generally less than 0.15. The bias its exclusion introduces is probably therefore minor unless it intrinsically has a very strong influence on the magnitude of the chronic illness variable.

So as to enhance the creditability of the dietary habits variable, FOOD, we quote from Survey Research Center (1972a, p.75):

> "Since expenditure on food is a relatively easy to measure proxy for adequate nutrition and is one of the study's more important variables, much care has been taken to improve the technique of asking these questions; several refinements, but no added objectives, have resulted in a few changes to these questions over the five waves of the survey."

Accepting the assertion that the amount of food expenditures was one of the most carefully treated questions in the entire SRC survey effort, the issue remains as to whether these expenditures, even when stated relative to food "needs," are capable of providing useful information on the etiology of illness. Certainly they can provide no information on dietary contributions to particular diseases unless expenditures on particular food groups are known. But then we are dealing in any case only with generalized measures of self-reported health status. As for the use of expenditures on food rather than actual food consumption, one's comprehension of this measure is aided if it is viewed as a proxy for a stock variable relating to the history of the individual's investments in diet. Real capital in the hospital industry is not measured in terms of gadgets and buildings but rather as the discounted value of the accumulated investments. Similarly, dietary adequacy may be measured as the discounted value of the individual's accumulated expenditures on food. FOOD, which is simply current expenditures on food relative to a "needs" standard, will generally tend to be positively related to this discounted value.

The intent of including the CHCH, RISK and RELG variables is to capture acquired behavioral traits consistent with an out-of-the-ordinary aversion to health-endangering activities. We hope at least some of those forms of health-enhancing everyday behavior not otherwise available in the data set collapse into these variables. Among these forms would be regulatory getting six to eight hours sleep, a tightly-knit and emotionally supportive family life, a healthy mix of foods consumed, and the many other lifestyle factors to which assorted medical commentators variously attribute the production and protection of good health.

INSR, a dummy variable referring to whether or not the individual is covered by medical insurance, should be correlated with the individual's consumption of medical care. The variable should be negatively related to the price of medical care that the individual faces and therefore positively related to the quality of medical care he has consumed. If medical care improves health or maintains good health, then the medical insurance variable should have a negative coefficient in both the acute and the chronic illness dose-response functions. Our use of this variable in a dose-response function might be criticized on grounds that it is serving as a proxy for the quantity of medical care consumed, where this quantity and the proxy are the consequence of current period decisions. We admit the possible validity of this view but nevertheless chose to retain INSR as our only available proxy likely to be strongly associated with the individual's adult history of medical service consumption. In short, we assume that the benefits to estimation from including a plausibly relevant variable (a history of the individual's adult consumption of medical services) outweigh the losses to estimation incurred by employing a current period decision variable as an explanatory variable in a single equation structure.

Among the environmental variables, all the air pollution variables, as well as DENS and CHEM, are expected to have positive signs for both acute and chronic illnesses. People who live in crowded conditions are in closer contact with other individuals, making personal sanitation more difficult, and increasing the probabilities of contracting whatever communicable illnesses plague others. The contacts of workers in the chemicals and metals manufacturing sectors are not so much with carriers of communicable illnesses, but rather with exposures to toxic substances in the work place. These exposures are thought to exceed those of the rest of the population.

Hippocrates, 460-337 B.C. (1939) and the writers of a large literature descending from those ancient times have asserted a sort of climatic determinism with respect to health.7/ We briefly acknowledge this literature by considering two climatic variables, COLD, to represent the extent of freezing weather, and ULTV, to indicate the amount of sunshine. Although the literature in this area says that climate has an influence on health, any advice it gives as to whether these climatic parameters are harmful or beneficial is unsettled. We therefore prefer not to make assertions about the signs to be expected for the coefficients of these variables.

A great many more variables for each of our variable classes is available on the SRC survey tapes. In addition, since the county of residence is known for each individual respondent for each year of the survey, additional environmental and general area information could be combined with the SRC tapes. Many more variables could be constructed from the basic SRC information. We did initially consider some other definitions and versions of the variables in Table 5.3, but this list should provide a reaonable description of the data we had available.

Before proceeding to the presentation and discussion of the dose-response functions, there are several salient characteristics of the constructed data set that do not necessarily have clear implications for the results but which nevertheless provide form and a setting for them. Tables 5.2 and 5.4 are thus worthy of some attention. The reader is reminded, however, that these tables are incomplete: they are only representative of the samples used to estimate the dose-response functions.

Note that three of the characteristics of Table 5.4 are consistent with a high proportion of the individuals in the sample having lived for long periods in one locale. People who live within walking distance of relatives, have always lived in one state, and have never moved to take a job elsewhere have likely had a long history of exposures to the outside air pollution of one municipality. In short, the SRC data allow one to compensate somewhat for the lack of a long data series on the pollution

Table 5.4

Proportions of Entire Survey Research Center Sample Processing a Particular Characteristic During 1971

Characteristic	Percent
Asset income $\leq$ \$500	81.1
Children $\leq$ 25 years in family unit	51.3
Has relatives living within walking distance	42.6
Employed head	72.7
Unemployed head	2.2
Retired head	16.6
Housewife head	6.7
Student head	1.6
Working wife	33.3
Disabled person in family other than head	3.8
Neighborhood of detached single-family homes or lesser density	65.9
Rents dwelling unit	37.8
Always lived in one state (1970 data)*	40.4
Never moved from a community for a job change (1970 data)*	57.9
Disabled head	21.8

---

*These proportions are not indicated in the code book describing the 1971 data. It is highly unlikely that they differ significantly from the 1971 proportion.

exposures of sample individuals. If one is willing to assume that relative pollution concentrations among locations have been reasonably constant over time, then he can at least loosely grasp the effects of cumulative exposures on differences in health states. These cumulative exposures might not be terribly relevant with respect to acute illness, but they can be highly important with respect to chronic illness. Therefore in all our empirical efforts dealing with chronic illness, we deal only with sample individuals who have always lived in one state or who have never moved for a job change. Even though this partitioning by no means guarantees that we fully capture the cumulative air pollution exposures of the sample individuals, we believe that it does so to a substantially greater degree than do most air pollution epidemiology data sets.

The proportion of sample individuals who profess disabilities consistently approximates one out of every five. Over the nine year interval of the data set, it ranges from a low 18.2 percent in 1974 to a high of 23.6 percent in 1969. In fact, only for the 1974 and 1975 entire SRC population samples was the proportion disabled below 20 percent (in 1975, the proportion was 18.4 percent). These lower proportions for 1974 and 1975 are probably due to the rather drastic drop in the mean age of the sample population occurring between 1973 and 1974, which is reflected in the mean values for the AGEH variable in Table 5.2. The drop causes the proportion of the SRC sample that reports being disabled to better approximate the proportion disabled in similar area probability samples of the U.S. civilian non-institutionalized population. These other samples generally tend to have ten to fifteen percent of their individuals suffering from self-reported disabilities.

A glance at Table 5.2 shows that the number of individuals employed in the chemicals and metals manufacturing sector is usually too small, given sample sizes of about 400, to estimate reliably the extent to which the exposures associated with this employment generate illness. As earlier noted, the 1973 SRC data include information on three-digit occupational codes by three-digit industry for the sample individuals. If, after having carefully perused the data to ascertain exactly which occupations in which industries involve substantial exposures to toxics, the entire SRC population sample were to be used to estimate an acute or chronic illness dose-response function, one might have sufficient degrees of freedom available to obtain reliable coefficients for these manufacturing sectors. At best, one or two of the samples we employ here have enough sample individuals employed in these sectors to be slightly suggestive about an association between exposures in them and acute or chronic illnesses.

Finally, when evaluating the empirical results reported in this study, one must face the question of the accuracy of respondent recall. Since there exists no data base referring to contemporaneously observed sample individual behavior and status, one's judgments about accuracy must necessarily be more-or-less personal and introspective. The following pair of facts can aid in the formation of this judgment. First, all respondent interviews were conducted within 12 months of the year for which respondent behavior and status was to be reported. Thus the longest interval that could pass between some respondent event and his reporting of that event was 23 months. In all years, however, the great bulk of the interviewing was completed by June of the year following the year that was to be reported. For these respondents, the greatest time lag that occurred between an event and its reporting was 17 months. The smallest lag that could occur, since interviewing started in early March of the year following the year to be reported, was two months.[8]

Perhaps more relevant to the recall issue than the question of lags is the incentive respondents had to make mental or written note of their behavior and status to ensure accurate answers when the appointed time for their interviews arrived. Several points relevant to this incentive issue can be made. First, as reinterviewing "waves" (this is the SRC's term) passed, those original respondents who were hostile to the interviewing process and purpose probably removed themselves from the sample. We speculate that those who voluntarily stayed in the sample possessed a substantial incentive for accurate recall. This implies that data from later years is perhaps more reliable than data from earlier years. Second, those families that did remain in the sample became more familiar with what would be asked them with each reinterviewing wave and would therefore take more care to make mental or written note of events so that they could be accurately reported. This too implies that data from later years tends to be more reliable. Third, the respondents were paid a small sum ($5.00 - $7.00) for participating in the interview. Finally, after having completed an interview, the respondent was left a postcard that he was asked to send to the SRC in early January of the following year. This card informed the SRC of the respondent's current address. Those who did and did not return the cards were sent a reminder and a postcard in January, along with a summary explanation of empirical results from the interviewing of the preceding year. All who returned the postcards, whether or not reminded, were rewarded with an additional payment of $5.00. The SRC does not report the proportion of those who returned postcards, but, given the reinterview rate, one can reasonably conclude, that the return rate must have been fairly high. We judge from this that respondent interest in the survey must have been substantial, resulting in an incentive to keep rather careful track of behavior and status.

Aside from the detail of its information, the SRC sample and its combination with the air pollution data contain little that is remarkable relative to other data sets that have been used in air pollution epidemiology. Judging from the general sociodemographic attributes depicted in Tables 5.2 and 5.4, the sample in spite of our disregard of the SRC sample weights, appears to be close to a random sample of the U.S. civilian non-institutionalized population. The high proportion of non-whites does, however, raise some doubt about its exact representativeness.[9] The increasingly better control of sulfur dioxide emissions is clearly registered in Table 5.2, although control of particulates and nitrogen dioxide appears not to have exhibited much improvement over the nine-year interval. Table 5.2, by its failure to show data for variables in some years that appear in other years, exhibits both changes in the SRC interview formats as well as our deletion of variables in expressions estimated for some years when they were not statistically significant in expressions estimated for samples drawn from other years.

5.3 Estimates of Dose-Response Rates for Acute and Chronic Illness

To place any credence in the estimates presented in this section, one must believe that stochastic factors play a role in dose-response functions. Stochastic disturbances may have a greater or a lesser part to play than systematic biological, physical, economic, or social influences, but they nevertheless have a part. If all influences were entirely deterministic, the statistical procedures employed here (as well as all of epidemiology) would be unnecessary and redundant: all one would have to do to ascertain the values of the influences is go to the laboratory and perform the relevant measures. In fact, single observations on the phenomena of interest would suffice: if the observations conformed to the proposition, one would accept the proposition for now. Otherwise, it would be rejected. Biomedical research employs both laboratory and human population studies (and several different variants within each of these general classifications) to come to grips, most often with less then iron firmness, with dose-response functions. The use of these approaches and their variants is an admission that the functions involve significant stochastic elements.

Reference is made to rates rather than functions in the subtitle of this section because the empirical results reported apply only to changes in measured illness for one-unit changes in the explanatory variables of interest at the mean values of these dependent and explanatory variables. These changes could properly describe an entire dose-response function if and only if that function were linear in the original variables. Throughout the estimation procedure, we have employed linear functions for an assortment of reasons, not the least of which is that there appears to be no strong analytical or empirical precedence for doing otherwise with the generalized measures of ill-health we are using. We don't know whether the air pollution dose-health response function is supposed to be increasing at an increasing or a decreasing rate over a given interval. A linear function is the best available compromise between these two possibilities. The linear form is easily interpreted at a glance and, furthermore, relative to other readily estimated forms such as the multiplicative, it does not attenuate the potential influence of observations having extreme values. In the absence of knowledge about the functional form of the relationship one is estimating, the use of multiplicative and similar forms effectively reduces the variation of the sample and thus will often allow one to explain a larger proportion of the variation in the (rescaled) sample. For purposes of the present study, since we lack prior knowledge of the form of the dose-response functions, we wish to provide the extremes of good and ill health, and pristine and filthy air, full rein. This reluctance to reduce the influence of outliers, when combined with our use of data on individual human being rather than group averages, means that we reduce, if not completely deny, our chances of explaining large proportions of the variation among our basic observational units in acute and chronic illness.

Tables 5.6a and 5.6b present estimates for household heads of dose-response relations for acute illness and Tables 5.7a and 5.7b do the same for chronic illness. So as to reduce the extent to which cumulative exposures to outdoor air pollutants are unaccounted for, all the estimated chronic illness expressions employ as basic units of observation only household heads who have always resided in one state. This restriction is imposed for all chronic illness estimates throughout the chapter.

Substantial care has been taken to assure that all explanatory variables have either always been outside the household head's domain of control or have been established by his actions prior to the period being considered. Thus, variables such as the head's age, where he grew up, his father's education, and past financial status, his sex and race, and the cold, air pollution, and the ultraviolet radiation to which he is exposed at a particular location are matters over which he never has and never will exercise anything but the most trivial influence. They are inalterable. Other variables such as the severity of any disabilities he has, and his education, marital status, and family size were certainly influenced by his decisions. However, the impact of past decisions on the current values of these variables will, for nearly all adults, overwhelm any potential impact of decisions made within any current 12 month period. The economic sector within which one is employed and the rooms per family members in one's housing are perhaps subject to more immediate control but, for the great bulk of people, are not very quickly or readily adjusted. Assertions of predetermination are clearly inaccurate for most of the life-style variables. One's current cigarette consumption, exercise, and dietary habits, etc., are quickly adapted to changing circumstances. Yet one might also reasonably argue that even these current adaptations are isomorphic to acquired habits, and can thus be employed as proxies for these predilections. In fact, for items such as medical insurance, food and cigarettes, there is abundant evidence in the empirical consumer demand literature that the quantities individuals consume are quite insensitive to price changes, at least for the range of price changes likely to occur in a year. Similarly, these habits tend to persist for some time in the face of substantial yearly income changes. Finally, introspection says that one's religious and risk aversion attitudes are the result of the accumulated experiences and learning of a lifetime rather than a momentary diversion that will serve only until a new fad comes to one's attention.

A rather large data set like the SRC survey, when joined with a quite sparse set of _a priori_ propositions with which to restrict the expressions to be estimated, leads one into temptation. In particular, using an unchanging set of sample observations, one is tempted to add and delete variables and try assorted functional forms until a result is obtained that, on statistical grounds alone, looks good; that is, the coefficients attached to the explanatory variables all have common sense or _a priori_ acceptable signs and are generally statistically significant at high levels. Moreover, summary statistics such as the coefficient of determination are high and standard errors of estimate are low. Quite frequently, the results of this "data-grubbing" are reported without any description of the manipulations lying behind them. As is well-known, this practice can introduce substantial biases into estimated coefficients. In the words of Selvin and Stuart (1966, p. 21):

> "... any preliminary search of data for a model, even when the alternatives are predesigned, affects the probability levels of

> all subsequent tests based on that model on the same data, and in no very simple way, and also affects the characteristics of subsequent estimation procedures. The only valid course is to use different data for testing the model dredged from the first set of data."

We have not conformed absolutely to this dictum, but have nevertheless followed it rather closely.[10]

In Tables 5.6a, 5.6b, 5.7b, each estimated expression is numbered, with each number in each table corresponding to an entirely new sample drawn at random from the entire SRC population sample or that portion of the SRC sample meeting certain imposed conditions. Thus, for example, in Table 5.6a expressions (1A), (1B), and (1C), are estimated from the same set of observations but the expressions (1) and the expressions (2) are estimated from entirely different samples. Since the availability of variables in the SRC data set can differ greatly from year-to-year, and the definitions of variables can differ slightly, it is not possible to exploit formal statistical tests for replication. Nevertheless, if the different samples do yield similar results for a particular set of variables, a dimension is added to the estimation procedure that undeniably adds information and confidence in the results.

Even though a modicum of something resembling data-grubbing is present in the estimation of expressions like (1A), (1B), and (1C) in Table 5.6a, it does not involve anything more than using the same data set to reestimate expressions in which nothing other than the air pollution variables has been changed. Thus, though (1) in Table 5.6a involves three expressions, only three "runs," with one run for each combination of air pollution variables, was performed.

Table 5.5 is a table of simple correlation coefficients for a representative sample. These coefficients, of course, differed from one sample to another, but the table provides a good idea of the general patterns of intercorrelation among the variables that were estimated by the various samples. As a glance at the table shows, there is very little linear association between the air pollution variables and any single other variable used to explain acute and chronic illness. No one of these other explanatory variables linearly accounts for more than 23 percent of the variance of an air pollution variable, and, in most cases, the variance accounted for is considerably below ten percent. Similarly, the intercorrelation among variables other than the air pollution variables tends to be very low. This, of course, does not mean that strong nonlinear associations between single variables are absent. Neither does it mean that close associations between the air pollution variables and linear or nonlinear combinations of other explanatory variables are not present. Although there exist some statistics that purport to test for these latter two possibilities, we have not employed them in this report. We thus proceed as if the fact that linear associations between single explanatory variables are typically low implies that multicollinearities among all explanatory variables (except for the air pollution variables) are unlikely to inflate the standard errors of coefficients, thereby causing certain

Table 5.5

Matrix of Simple Correlation Coefficients for a 1971 Representative Dose-Response Function Sample

	LDSA	EDUC	CIGE	EXER	FOOD	RISK	AGEH	DSAB	FMSZ	SEXH	INSR	SVGS	CHEM	CITY
LDSA	1.000	-0.139	-0.112	-0.075	-0.202	-0.132	0.201	0.700	0.037	-0.117	-0.301	-0.175	0.029	-0.008
EDUC		1.000	0.001	0.230	0.442	0.454	-0.155	-0.153	-0.191	0.217	0.401	0.354	0.201	0.057
CIGE				0.006	0.077	-0.268	-0.091	-0.062	0.126	0.303	0.186	0.018	-0.037	0.303
EXER					0.054	0.134	-0.106	-0.059	0.038	0.167	0.151	-0.150	-0.026	-0.025
FOOD						0.434	0.157	-0.172	-0.369	0.311	0.414	0.501	0.059	0.012
RISK							-0.021	-0.200	0.173	0.427	0.678	0.043	-0.032	0.068
AGEH								0.231	-0.081	0.018	0.001	0.284	0.076	-0.094
DSAB									-0.035	-0.184	-0.325	-0.165	-0.024	-0.048
FMSZ										0.081	-0.206	-0.188	-0.004	0.047
SEXH											0.217	0.215	0.033	0.088
INSR												0.418	0.033	-0.027
SVGS													0.065	-0.058
CHEM														0.035
POOR	0.156	-0.137	-0.124	0.024	-0.239	-0.133	0.115	0.137	0.108	0.030	-0.060	-0.165	-0.052	0.076
FEDU	-0.164	0.188	0.137	0.169	0.501	0.214	-0.176	-0.135	-0.182	0.147	0.177	0.119	-0.012	0.100
RACE	-0.160	0.419	0.306	0.170	0.732	0.447	0.180	-0.136	-0.235	0.353	0.413	0.414	0.050	-0.017
ACUT	0.006	-0.013	0.023	-0.080	0.289	-0.053	0.026	0.155	-0.059	-0.071	0.046	-0.024	-0.018	-0.055
TSPT	0.096	-0.012	-0.084	-0.084	0.044	0.042	-0.085	-0.043	0.034	0.110	0.066	0.155	0.091	0.042
TSPN	0.153	-0.058	0.079	-0.106	0.050	0.136	-0.097	0.080	0.055	0.091	0.069	0.231	0.100	0.074
TSPM	0.124	-0.024	0.079	-0.119	0.016	0.066	-0.086	0.150	0.040	0.133	0.047	0.172	0.083	0.056
SULT	-0.002	0.053	0.057	-0.126	0.018	0.069	-0.031	0.066	0.021	0.156	-0.054	0.108	0.015	0.072
SULN	-0.006	0.113	0.087	-0.129	0.072	0.136	-0.085	0.043	0.025	0.097	0.042	0.217	0.073	0.068
SULM	0.002	-0.021	0.045	-0.142	-0.009	0.058	-0.044	0.045	0.013	0.169	-0.031	0.119	0.031	0.071

	FEDU	RACE	ACUT	TSPT	TSPN	TSPM	SULT	SULN	SULM
POOR	-0.285	-0.130	-0.094	0.054	0.067	0.056	0.004	0.038	0.005
FEDU		0.326	0.030	0.094	0.039	0.075	0.039	0.052	0.065
RACE				-0.071	0.008	0.056	-0.012	-0.082	-0.100
ACUT				0.155	0.074	0.036	0.119	0.122	0.122
TSPT					0.922	0.970	0.441	0.742	0.658
TSPN						0.976	0.652	0.897	0.821
TSPM							0.622	0.861	0.800
SULT								0.868	0.945
SULN									0.939

Table 5.6a

Dose-Response Rates for ACUT: Unpartitioned Samples

Year	(1A)		(1B)		(1C)		(2)		(3)	
	1967		1967		1967		1968		1969	
Variable	β	s	β	s	β	s	β	s	β	s
DSAB	20.541	5.862	21.140*	5.947	21.520*	5.854				
LDSA							47.04*	16.08	3.252	12.290
AGEH	-2.486	1.650	-2.068	1.246	-1.895	1.637	-1.306	1.456	-1.208	1.097
EDUC	4.086	13.344	4.155	13.540	4.462	13.300				
MARR	-12.561	81.952	-8.362	81.660	-21.280	80.500	16.610	35.560	0.065	29.11
POOR	-24.264	40.419	-24.060	40.800	-26.120	40.180	-29.80	34.03	-52.320*	27.030
RACE	-87.746*	46.328	-95.090*	49.95	-109.900*	53.220				
SEXH	-17.564	87.082	-7.666	86.450	-20.370	85.480				
EXER									-66.732*	34.930
FOOD	-0.062*	0.039	-0.063*	0.033	-0.066*	0.037	-0.056*	0.023	-0.071	0.175
NCIG	17.943	11.801	18.520	12.010	20.170*	11.760	16.130*	9.844		
RELG	-12.561	81.958								
RISK	-9.392	16.670	-11.170	17.190	-13.770	16.720	-17.676*	12.330	-25.960*	9.668
INSR	20.84	59.05	15.150	59.31	13.380	60.380	88.710**	47.090	67.510**	37.420
CHEM										
DENS									1.127	7.429
NOXT										
NOXM										
NOXN										
SULT	1.857*	1.033								
SULM			1.488*	0.733			1.518*	0.925	-1.199	0.951
SULN					0.722*	0.372				
TSPT	-0.432	0.681								
TSPM			-0.442	0.648			-0.963	0.606	1.122*	0.765
TSPN					-0.120	0.261				

(continued)

Table 5.6a
(continued)

Year	(1A)	(1B)	(1C)	(2)	(3)
Variable	β 1967 s	β 1967 s	β 1967 s	β 1968 s	β 1969 s
Constant	410.960	322.546	320.309	447.874	283.201
R^2	0.307	0.296	0.313	0.175	0.182
S.E.	164.745	166.030	164.108	317.210	264.023
F	(13,80) = 4.731	(13,80) = 4.594	(13,80) = 4.800	(10,389) = 6.139	(10,389) = 5.473
η_{NOX}					
η_{SUL}	0.308	0.353	0.544	0.326	
η_{TSP}					0.474

(continued)

Table 5.6a
(continued)

Year	(4)		(5)		(6)		(7)	
Variable	β 1970	s	β 1971	s	β 1972	s	β 1973	s
DSAB					180.800*	26.550		
LDSA	76.490	13.920	47.990*	14.590			19.340	14.740
AGEH	0.485	1.153	2.542*	1.199	1.411	1.563	0.355	1.051
EDUC			-15.800*	8.370			2.550	7.818
POOR	-63.200*	27.360	-49.260*	26.640	128.100*	41.550	-14.190	28.110
RACE					4.435	22.680	-36.450*	20.460
SEXH	18.490	31.070	-85.170*	31.090	20.740	43.290	-123.600*	29.390
CIGE								
EXER	-48.620	30.260	-30.150	31.560				
FOOD			-0.030	0.021				
NCIG								
RELG								
RISK	-37.120*	9.872	-21.796*	9.876				
INSR	9.885	36.780	7.439	40.330				
CHEM	-3.280	55.020			-99.730	110.300	54.410	76.740
DENS					3.529*	1.597	0.056	1.075
NOXT								
NOXM								
NOXN					-0.972	0.684	0.223*	0.124
SULT								
SULM	2.520*	1.104	2.257	2.259			-0.361	3.305
SULN								
TSPT								
TSPM			1.453*	0.764	1.782*	0.780	-0.249	0.314
TSPN								

(continued)

Table 5.6a
(continued)

Year	(4)	(5)	(6)	(7)
Variable	β 1970 s	β 1971 s	β 1972 s	β 1973 s
Constant	305.260	172.464	-78.317	175.040
R^2	0.123	0.123	0.169	0.095
S.E.	262.333	252.936	394.533	254.413
F	(9,390) = 6.104	(11,388) = 4.926	(9,390) = 8.836	(11,388) = 4.435
η_{NOX}				0.618
η_{SUL}	0.361	0.518		
η_{TSP}			0.497	

*Significant at the 0.05 level of the one-tailed t-test.

**Significant at the 0.05 level of the two-tailed t-test.

Table 5.6b

Dose-Response Rates for ACUT: Partitioned Samples

Year	(1) 1967 Always lived in 1 state		(2) 1969 RINC = ≤ $7,500		(3) 1969 3 ≤ NCIG ≤ 6		(4) 1970 1 ≤ DSAB ≤ 3	
Variable	β	s	β	s	β	s	β	s
DSAB	42.056*	6.538						
LDSA			111.200*	37.590	17.960	13.810	-94.990*	34.430
AGEH	0.384*	0.187	2.889	3.488	-2.383	1.435	1.215	
EDUC	2.716	2.302						
MARR	-17.037	85.185	117.800	75.310	35.290	38.710	56.630	99.450
POOR	31.832	46.812	-116.400	78.780	5.323	31.210	-66.900	86.310
RACE	-60.549	53.583						
SEXH	13.327	9.005						
FOOD	-0.061	0.057	-1.648*	0.559	-0.084	0.218	-0.168	0.519
NCIG	5.643	3.239*	36.940*	22.710	34.030*	18.620		
RISK	-4.047	17.600	32.950	27.840	-4.700	12.380	1.938	35.750
INSR	-75.286	70.361	80.820	86.440	-71.390*	42.490	-54.560	117.000
SULM	-0.992	7.631	5.135*	3.020	0.007	0.831	0.114	2.930
TSPM	1.765*	0.865	-4.031	3.020	-0.594	0.480	1.215	2.210
EXER							200.600*	125.600
DENS							-31.710	21.700
Constant	121.290		566.723		165.600		482.897	
R^2	0.152		0.186		0.076		0.122	
S.E.	352.420		443.738		243.090		449.633	
F	(14,306) = 6.621		(10,150) = 3.431		(10,268) = 2.191		(10,114) = 1.585	
NSUL			0.565					
NTSP	0.952							

*Significant at the 0.05 level of the one-tailed t-test.

**Significant at the 0.05 level of the two-tailed t-test.

Table 5.7a

Dose-Response Rates for LDSA: Unpartitioned Samples[a]

Year	(1)		(2)		(3a)		(3b)		(5A)	
	1967		1968		1970		1970		1971	
Variable	β	s	β	s	β	s	β	s	β	s
DSAB			3.286*	0.227	0.554**	0.035	0.550**	0.035	0.808**	0.049
AGEH	0.003	0.007	-0.002	0.007	0.005	0.004	0.005	0.004	0.007*	0.004
CITY										
EDUC	0.079	0.054	0.170	0.416	0.013	0.029	0.001	0.029	-0.057	0.030
FEDU					-0.044	0.037	-0.043	0.037	-0.044	0.035
MARR	0.204	0.284								
POOR	0.188	0.157	0.135	0.163	-0.069	0.103	-0.065	0.103	0.086	0.096
RACE	0.344	0.200			0.072	0.488	0.088	0.487	-0.057	0.119
SEXH	0.410	0.297			0.139	0.114	0.132	0.113	0.233**	0.109
FOOD	-0.7×10^{-4}	0.26×10^{-3}	0.002*	0.001	-0.902	0.975	-0.924	0.973	-0.13×10^{-3}	0.81×10^{-4}
NCIG	0.023	0.047	-0.089*	0.041						
RISK	-0.009	0.006								
INSR	-0.152	0.245	-0.336*	0.218	-0.454*	0.129	-0.459*	0.129	-0.496*	0.125
CHEM					-1.645**	0.575	-0.097	0.575	-0.002	0.916
NOXT										
NOXM										
NOXN										
SULM	0.0036	0.0025	0.0067*	0.0035						
SULN										
TSPM	0.0021	0.0037	-0.0036	0.0024	0.0028*	0.0011			0.0019	0.0017
TSPN							0.0018*	0.66×10^{-3}		
Constant	-0.636		0.631		2.980		2.924		0.265	
R^2	0.094		0.371		0.525		0.526		0.530	
S.E.	0.835		0.736		0.964		0.963		0.904	
F	(12,134) = 2.158		(9,390) = 25.580		(11,388) = 38.920		(11,388) = 39.170		(11,388) = 39.800	
η_{NOX}										
η_{SUL}										
η_{TSP}					0.278		0.341		0.268	

(continued)

Table 5.7a

(continued)

Year	(5B)		(6A)		(6B)	
Variable	β 1971	s	β 1972	s	β 1972	s
DSAB	0.809**	0.049				
AGEH	0.007*	0.004	0.023*	0.005	0.020*	0.005
CITY			-0.057	0.050	-0.085	0.045
EDUC	-0.058	0.030	-0.081	0.045	-0.125**	0.043
FEDU	-0.043	0.035	0.007	0.050	0.048	0.055
MARR						
POOR	0.088	0.097	0.104	0.147	0.116	0.145
RACE	-0.054	0.119	-0.272	0.160	-0.220	0.154
SEXH	0.240**	0.109	-0.156	0.141	-0.182	0.142
FOOD	$-0.13\text{x}10^{-3}$	$0.81\text{x}10^{-4}$				
NCIG						
RISK						
INSR	-0.499*	0.125				
CHEM	-0.016	0.917				
NOXT						
NOXM			-0.0007	0.0017		
NOXN					$0.14\text{x}10^{-3}$	$0.89\text{x}10^{-3}$
SULM						
SULN						
TSPM			0.0028	0.0027		
TSPN	$0.59\text{x}10^{-3}$	$0.73\text{x}10^{-3}$			0.0030*	0.0013
Constant	0.181		0.701		1.054	
R^2	0.529		0.119		0.134	
S.E.	0.905		1.347		1.315	
F	(11,388) = 39.680		(9,390) = 5.879		(9,390) = 6.706	
η_{NOX}						
η_{SUL}						
η_{TSP}			0.376		0.630	

(continued)

Table 5.7a
(continued)

Year	(7A)		(7B)		(8A)		(8B)	
Variable	β 1973	s	β 1973	s	β 1974	s	β 1974	s
DSAB								
AGEH	0.017*	0.044	0.017*	0.004	0.017*	0.005	0.017*	0.005
CITY	0.155	0.140	0.180	0.140				
EDUC	-0.122*	0.043	-0.128*	0.043	-0.118**	0.049	-0.111**	0.049
FEDU	0.059	0.052	0.057	0.053				
MARR					-0.291	0.221	-0.287	0.221
POOR	0.050	0.143	0.029	0.142	0.300*	0.151	0.305*	0.151
RACE	-0.208	0.154	-0.202	0.155				
SEXH	-0.207	0.141	-0.209	0.141	-0.001	0.230	-0.001	0.230
FOOD					-0.060	0.066	-0.067	0.067
NCIG								
RELG					-0.459	0.284	-0.457	0.284
INSR								
CHEM					-0.161	0.327	-0.131	0.325
NOXT								
NOXM					0.0023	0.0017		
NOXN							0.0046*	0.0025
SULM	0.0033	0.0037			-0.0047	0.0062		
SULN			0.003	0.002			0.0002	0.0022
TSPM	0.0017	0.0015			0.0008	0.0028		
TSPN			0.0004	0.0017			-0.0007	0.0019
Constant	0.309		0.505		-0.687		-0.828	
R^2	0.106		0.109		0.118		0.112	
S.E.	1.303		1.303		0.966		0.964	
F	(9,390) = 5.785		(9,390) = 5.290		(11,214) = 4.591		(11,214) = 4.693	
η_{NOX}					0.363		1.143	
η_{SUL}								
η_{TSP}								

(continued)

Table 5.7a
(continued)

*Significant at the 0.05 level of the one-tailed t-test.

**Significant at the 0.05 level of the two-tailed t-test.

[a] All observations in this table are limited to individuals who have always lived in one state.

Table 5.7b

Dose-Response Rates for LDSA: Partitioned Samples[a]

Year	(1) 1971 50-cities		(2) 1969		(3A) 1972 AGEH $\geq$ 45		(3B) 1972 AGEH $\geq$ 45		(4A) 1972 AGEH $\geq$ 45 & MILE $\leq$ 15		(4B) 1972 AGEH $\leq$ 45 & MILE $\leq$ 15	
Variable	β	s	β	s	β	s	β	s	β	s	β	s
DSAB	-0.168**	0.068	2.462*	0.108					0.904**	0.054	0.897**	0.057
AGEH	0.025*	0.006			0.028*	0.008	0.029*	0.008	0.020*	0.006	0.021*	0.006
CITY	-0.401**	0.190			-0.007	0.055	0.031	0.059	-0.115**	0.046	-0.022	0.043
EDUC	-0.057	0.047			-0.080	0.048	-0.124**	0.045	-0.055	0.035	-0.045	0.037
FEDU	-0.048	0.040			0.001	0.060	-0.062	0.060	-0.035	0.046	0.020	0.047
MARR												
POOR			0.163*	0.067	0.161	0.155	0.112	0.151	0.161	0.155	0.078	0.121
RACE					-0.268	0.166	-0.028	0.169	0.078	0.130	-0.272*	0.129
SEXH					-0.178	0.153	-0.217	0.156	-0.119	0.119	-0.111	0.120
FOOD	0.015	0.019	0.12×10^{-3}	0.48×10^{-3}								
NCIG	0.064	0.045										
RELG	0.050	0.679										
RISK			0.028	0.025								
INSR			-0.012	0.087								
CHCH			-0.005	0.067								
COLD	0.001	0.002										
NOXT					0.0021*	0.0013						
NOXM	0.0047*	0.0023					0.0017	0.0012	0.0021*	0.0012		
NOXN											0.0013*	0.0007
SULT												
SULM	-0.0018	0.0032	0.0025*	0.0013								
SULN												
TSPT					-0.0008	0.0035						
TSPM	0.0078*	0.0038	0.85×10^{-3}	0.20×10^{-2}								
TSPN											-0.0009	0.0011
ULTV	-0.51×10^{-5}	0.16×10^{-3}										
Constant	-1.018		0.005		-0.378		-0.037		-0.285		0.005	
R^2	0.210		0.624		0.078		0.083		0.464		0.439	
S.E.	1.563		1.265		1.435		1.438		1.101		1.120	
F	(14,304) = 5.762		(9,340) = 62.58		(9,390) = 5.899		(9,390) = 5.899		(10,389) = 33.630		(10,389) = 30.490	
η_{NOX}	0.470		0.608		0.258		0.301		0.369		0.327	
η_{SUL}			0.514									
η_{TSP}	0.948											

*Significant at the 0.05 level of the one-tailed t-test.

**Significant at the 0.05 level of the two-tailed t-test.

[a] All observations in this table are limited to individuals who have always lived in one state, except for the observations in (2). These are limited to individuals who currently live within walking distance of relatives.

[b] The air pollution variables for this expression refer to arithmetic mean 1969-71 geometric mean concentrations in $\mu g/m^3$. The referenced 50 cities are 50 of the 60 cities used in the aggregate mortality study that form a part of this report.

coefficients to appear statistically non-significant when they are properly viewed as significant.

There are, however, two very important exceptions to the supposed absence of a multicollinearity problem: the types of air pollution tend to be very highly correlated and different moments of the same pollutant also are closely associated. As Table 5.5 shows, total suspended particulates and sulfur dioxide appear to have a very high linear association as do all the moments of a particular air pollutant. If one were to introduce nitrogen dioxide in Table 5.5, the linear association between this pollutant and total suspended particulates and/or sulfur dioxide would also be large, though somewhat smaller than that between the latter two pollutants. For example, in 1975, the simple correlation coefficient between various measures of total suspended particulates and nitrogen dioxide is never less than 0.50 and sometimes reaches into the 0.70 or greater range. Given these close linear associations among the types of air pollution, we are reluctant to assign a health effect to a particular pollutant. Instead, it seems preferable to make the assignment to the outdoor air pollution phenomenon. In addition, when one or more air pollutants appear as explanatory variables in an estimated dose-response expression, the standard errors of each will tend to be somewhat inflated. Thus, a few of the air pollution coefficients to which we do not attach significance sometimes would be significant if one or more of the other air pollution variables were removed. Similarly, some of those air pollution coefficients that are significant would be more significant with the removal of a companion variable from the expression.

The above discussion does not deal with a dilemma posed by the issues of bias and multicollinearity. If the different types or moments of air pollution have separable impacts on health, then one biases the coefficients of the remaining explanatory variables by deleting one or more of the air pollution variables. Nevertheless, if one includes the highly collinear air pollution variables, he reduces the apparent statistical significance of any one of them. In this study, we do not directly attack the dilemma by constructing and then applying rigorous criteria for choice. We choose an easier and less rigorous course by estimating some expressions, each from a different sample, that include all the types of air pollution, while including only one type of air pollution in other expressions. To a very substantial extent, this course was forced upon us by circumstances: for some years over the nine-year SRC survey interval, there was no available information on particular types and moments of the air pollution variables.

Table 5.5 exhibits one other intercorrelation that is a cause for concern, namely a simple correlation coefficient of 0.70 between LDSA and DSAB, i.e., between the duration of a chronic illness and its self-reported severity. Relative to most other samples of the study, this intercorrelation is a bit low. For most samples, it is closer to or in excess of 0.80. Certainly, the length of a disease and its severity are not identical. In fact, one might expect those who are severely disabled to have relatively short disease durations: they are more likely to die. We may thus have increased the intercorrelation between these two variables by not making DSAB be monotonically increasing. The high intercorrelation arouses suspicions about whether the two variables might be measuring the same thing, a clearly ridiculous state, if one is trying to explain the covariation between the two variables. Furthermore, if air pollution is expected to lengthen the duration of an illness, there is obvious reason to think that it will also make an illness more severe. More accurately perhaps, air pollution causes illness and increases the severity of preexisting illness, thus in a recursive fashion lengthening, for those who survive, illness duration. This implies that the estimated expressions which include DSAB as an explanatory variable are actually reduced form expressions, where DSAB is determined within the structural system. As a result, the single equation estimates with DSAB as an explanatory variable are not asymptotically efficient although they are consistent since DSAB is the only explanatory variable that would be determined within the structure of a recursive system. If instead of DSAB being a determinant of LDSA, it is actually another measure of the same thing in respondents' views, then DSAB must be dropped from the estimated expression. For the expressions estimated from some samples we include DSAB; for other samples, we delete it, using whichever of the preceding rationales conforms to the estimated expression. As we will see, inclusion or exclusion doesn't really make much difference to the signs and magnitudes of the coefficients for the major variables of interest, the air pollution variables.[11/]

In estimating dose-response expressions for chronic illness, we have used LDSA rather than (or in addition to) DSAB because only the former is stated in cardinal terms. LDSA, however, retains one disadvantage of DSAB; as presented on the SRC tapes, it takes on only five values. Although the first four of these values apply to approximate two-year intervals, the last value might better be termed "a long time," since it is meant to apply to disabilities lasting eight or more years. If one interprets, as we shall do in this chapter, this last value to be equal to exactly ten years, then the dependent variable for chronic illness has a measurement error that biases it downward, causing the effects of the explanatory variables to be underestimated. This could be a serious source of error since about 40 percent of those who are disabled in any given SRC survey year, or seven to eight percent of the total SRC respondent population, profess to have been disabled for eight or more years. Given this problem, which we disregard until a succeeding section, it is perhaps preferable to interpret the coefficients attached to the explanatory variables in the estimated chronic illness dose-response expressions as the proportion of one of the discrete values comprising LDSA associated with a one unit change in the relevant explanatory variable.

Yet another estimation issue is caused by the five discrete values assumed by LDSA. This small number of discrete values means that heteroskedasticity could be present in those expressions estimated by ordinary-least-squares techniques. Ideally, multinomial logit estimation would be employed; but because the number of parameters with multinomial logit estimation increases so dramatically when the dependent variable assumes more than two values, there is an explicit tradeoff between the misspecification possibly introduced by the use of ordinary-least-squares and the vastly increased cost and complexity of multinomial logit estimation. We have opted here for simplicity and lesser cost while not dismissing the heteroskedadicity issue: we estimate the chronic illness dose-response functions by ordinary-least-squares but peruse the estimated results by simple graphic techniques to check for the presence of heteroskedasticity. Even if this undesirable property is present, it does not follow that our estimates will be biased and inconsistent. They will not be efficient (they will not have the smallest variance in a class of unbiased estimators), but they will be unbiased and consistent. The problem with heteroskedasdicity is thus not with the estimated coefficients themselves but rather with the calculated standard errors. These standard errors are biased, thus invalidating the tests of significance for the estimated coefficients.

There are a number of results for acute illness in Tables 5.6a and 5.6b worthy of explicit note:

1) Of the seven different unpartitioned samples used to estimate acute illness dose-response expressions, statistically significant air pollution coefficients occur in all of them. Thus, an additional unit of air pollution, as defined by any of the variety of measures employed here, was associated with an increase of from one to four hours in average annual hours of acute illness. Except for 1973, magnitudes of the air pollution coefficients are quite stable from one sample to another, even though the specifications for the expressions often differ substantially. No tests have been performed to establish whether there are statistically significant differences in the air pollution coefficients across samples.

2) For the unpartitioned samples, the elasticity, η, of acute illness with respect to any of the air pollution variables (a unitless measure of the response of acute illness to variations in air pollution) is substantially less than unity. This implies that in the immediate neighborhoods of the sample values of these variables, average annual hours of acute illness is increasing at a decreasing rate with respect to increases in air pollution.

3) Two of the four partitioned samples in Table 5.6b do not have statistically significant air pollution coefficients. If air pollution has any impact upon the frequency of acute illnesses among individuals who are chronically disabled and who live in families where a pack or more of cigarettes is smoked, the estimataion techniques and sample sizes employed here are incapable of capturing it.

4) When measures of total suspended particulates and sulfur dioxide are included as explanatory variables in the same expression, the coefficient for them usually assumes a negative sign. Generally, total suspended particulates will take on the negative sign. Similarly, when sulfur dioxide and nitrogen dioxide are included as explanatory variables in the same expression, nitrogen dioxide often assumes a negative sign. For estimated expressions in which total suspended particulates and/or nitrogen dioxide are used as explanatory variables, but which do not include sulfur dioxide, both of the former air pollutants have positive signs. These sign switches could be due to the high linear associations among the pollution variables.

5) With some exceptions, an increase of one discrete value in either of the measures of chronic illness tends to increase the average annual hours of acute illness by from 20 to 40 hours.

6) With the sole exception of the variables for a poor childhood and race, the variables representing biological and social endowments fail to play a statistically significant and consistent role in the acute illness dose-response expressions. It is possible, of course, and perhaps even likely, that the race and childhood background variables are capturing many of the effects of low education, etc.

7) The life-style variables in the acute illness dose-response expressions consistently have the expected signs and are often statistically significant. This is particularly true for the exercise and nutritional adequacy variables: they reduce average annual hours of acute illness.

8) Contrary to expectations, the explanatory variable for availability of medical care, INSR, usually has a positive sign, implying that people with better access to medical care have more acute illness. We have no explanation for this other than a pure speculation that people with better access to medical care are more likely to recognize the symptoms of acute illness, perhaps because physicicians provide them with the information to recognize these symptoms. On the other hand, INSR might simply be a poor measure of the respondents' access to medical care.

9) Other than air pollution, only two alternative measures of the respondents' environments were employed as explanatory variables. These variables were used in only a limited number of samples. DENS, the number of persons per room in the respondents' residence, increased average annual hours of illness by more than three in the single sample where it was statistically significant. The variable for employment in the chemicals and metals manufacturing sectors had too small a number of individuals in each sample to yield statistically meaningful results.

10) Visual inspection of the residuals for expression (1A) of Table 5.6a and expression (1) of Table 5.6b did not reveal any serious heteroskedasticity problems.

We tentatively conclude from the preceding findings that the life-style and environmental variables, including air pollution, we have used probably play a significant role in acute illness. The evidence for the biological and social background and the access to medical care variables is substantially less clear both because of measurement problems and because racial differences in educational and childhood environment may be reflected in simple binary variables for race and a poor childhood. Finally, it should be noted that none of our expressions "explains" a very large portion of the variation in acute illness. The coefficients of determination never exceed 0.31 and are often about 0.10. Moreover, the constant term in each expression nearly always exceeds the sum of the coefficients of the explanatory variables. This is, of course, partly due to the scaling of the variables, but, given the number of binary variables (MARR, POOR, RACE SEXH, RELG, INSR), one might reasonably have not expected quite such a difference. The relatively unimportant role that many of the most statistically significant variables play in total variation in annual hours of acute illness is evident in the following partial coefficients of determination for variables appearing in various expressions of Table 5.6a: for expression (7), NOXM = 0.004, SEXH = 0.044, and and LDSA = 0.004; for expression (1B), SULM = 0.021, FOOD = 0.002, RACE = 0.043, NCIG = 0.029, and DSAB = 0.136; and for expression (5), TSPM = 0.013, POOR = 0.024, DSAB = 0.124. With no more than one or two exceptions, the two variables for chronic illness, LDSA and DSAB, made the largest contributions to explaining variations in annual hours of acute illness.

Tables 5.7a and 5.7b give the estimated dose-response expressions for chronic illness. The following features stand out in these expressions.

1. Of the twelve different partitioned and unpartitioned samples present in Tables 5.7a and 5.7b, air pollution coefficients are statistically significant in nine of them. Not all air pollution coefficients are statistically significant in the samples where more than a single air pollution variable appears, nor are the signs always positive for those air pollution coefficients that are statistically nonsignificant. No pattern similar to the negative signs that are attached to sulfur dioxide or other pollutants when sulfur dioxide is used as an explanatory variable in the acute illness dose-response expressions appears here, however. Of the samples having no air pollution coefficients statistically significant at the 0.05 level or better of the one-tailed t-test [expressions (1), (5), and (7) in Table 5.7a], all had air pollution coefficients with positive signs and t-values in excess of 1.0. Two of these samples [expressions (1) and (7)] had air pollution coefficients statistically significant at the 0.10 level of the one-tailed t-test. The magnitude (and signs) of the air pollution coefficients for expressions (1), (5), and (7) were similar to the magnitudes and signs of the air pollution coefficients for the other samples. They ranged between slightly less than 0.0020 and slightly more than 0.0045, with the bulk being between 0.0020 and 0.0030. This means that a change between 0.2 and 0.4 or 0.5 percent in one of the discrete values comprising LDSA is caused by a one-unit change in air pollution. In elasticity terms, these discrete values (index) of LDSA appear to be relatively unresponsive to changes in air pollution. Nearly all the elasticities of the discrete chronic illness index with respect to air pollution are in the 0.2 to 0.5

range, implying that a one percent change in air pollution generally causes a change in the index of substantially less than one-half of one percent. As was true for the acute illness dose-response expressions, this means that, in the immediate neighborhoods of the chronic illness index values and the air pollution values present in these samples, chronic illness duration increases at a decreasing rate with respect to increasing air pollution.

2. As earlier noted, translating the coefficients for the explanatory variables in the chronic illness dose-response expressions is invalid because the highest value in the index could, in real-time terms, be anything equal to or in excess of eight years. Nevertheless, if one assumes that the real-time involved in this last index value is equivalent to that in all the lower values, than the translation can be performed. With this assumption, the air pollution coefficients imply that an additional unit of air pollution is, on average, associated with an increase of from one and one-half to three and one-half days in the duration of chronic illness. As before, even with the aforementioned assumption, this rate is applicable only in the immediate neighborhoods of the chronic illness index values and the air pollution values present in the samples.

3. In those unpartitioned expressions where it is employed as an explanatory variable, the severity of the respondent's disabilities has a highly significant, positive, and strong effect on the duration of these disabilities. The partial coefficient of determination of DSAB with respect to LDSA was consistently about 0.40. The inclusion of DSAB in expressions did not appear to have an effect on either the magnitudes or the significances of the air pollution coefficients. Similarly, its presence or absence did not seem to make much difference to coefficients for the other explanatory variables.

4. Results for the biological and social endowment variables are mixed. Only respondent age is consistently significant with the expected sign. Generally, as expected, the level of the respondent's education is associated with lesser durations of chronic illness, but it is only occasionally significant. Poor parents tend to be consistently associated with increased chronic illness durations, but POOR is statistically significant in only one sample. Otherwise, variables such as CITY, FEDU, MARR, RACE, and SEXH contributed very little to the expressions. Rarely were they significant statistically. More importantly, their magnitudes and their signs proved to be expremely sensitive to whatever specification was adopted.

5. Because it is not clear that the magnitudes of lifestyle variables are independent of the duration of chronic illness, fewer of them were used, and those that were used were used less frequently, than in the acute illness dose-response expressions. EXER is an obvious case and it has not entered the chronic illness expressions. In fact, except for RELG, food adequacy is the only explanatory variable that enters the expressions estimated for more than one sample. It always has the expected sign but is never quite statistically significant at the 0.05 level selected for this study. On the rare occasions when they appear, both cigarette consumption and fundamentalist religious affiliations have the expected signs. RELG in expressions (7A) and (7B) just barely misses being crowned with statistical

respectability. Since religious affiliations seem likely to remain unchanged whether or not one is disabled, this variable probably should have been included for the expressions estimated from each sample.

6. The variable representing the availability of medical care, INSR, performed well for those four samples where it was used. Its sign was consistent with an interpretation that medical care availability reduces the duration of chronic illness. Unfortunately, its sign is also consistent with another interpretation: those who are chronically ill have difficulty procuring medical insurance. This latter interpretation means that INSR could be a function of LDSA. Given these conflicting interpretations, and having no information on which interpretation is likely to dominate, we have compromised and included INSR in some expressions while neglecting it in others. Its inclusion or exclusion does not appear to have any discernable effects on the coefficients for the air pollution variables.

7. Of the environmental variables, only CHEM seems worthy of comment. In the one expression where they appear, neither COLD nor ULTV were statistically signficant although COLD did have a positive sign. The statistical significance of CHEM in expression (3) of Table 5.7a should be disregarded. Expression (3) was estimated from a sample having only three people employed in the chemicals and metals manufacturing sector. None of these three people had a chronic disability.

8. With the exception of DSAB, none of the included explanatory variables explain substantial proportions of the variation in the index for duration of chronic illness. The air pollution variables, taken together, explain no more than two percent of the variation in LDSA; AGEH sometimes explains as much as five percent and EDUC usually explains around three percent of this variation. As with the acute illness dose-response functions, we have not been able to account for very much of the sources of variation in chronic illness.

9. Table 5.7b exhibits the estimated expressions for samples that were restricted to the values of the variables indicated at the top of each column. Contrary to similar restrictions placed on the samples for the acute illness expressions, these restrictions did not alter the explanatory variable coefficients in any noticeable fashion.

10. The patterns of the residuals for several of the expressions in Table 5.7a have been visually inspected for evidence of heteroskedasticity. When this problem is present, it appears that the residuals tend to increase with increasing values of the dependent variable. Because the highest discrete value of LDSA has no upper bound, it is likely that the true variance of the sample tends to increase with increasing values of LDSA. As Kmenta (1971, p. 256) shows for expressions with a single explanatory variable, if the residuals and the sample variance are positively associated, the standard errors of the coefficients for the explanatory variable will be biased downward, causing the t-value to be too great. This need not be true, however, for expressions with multiple explanatory variables. The extent to which this has resulted in exaggeration or underestimates of the levels of significance for the chronic illness dose-response expressions is presently

Table 5.8

Lagged Effects of Total Suspended Particulates upon Duration of Chronic Illnesses (LDSA) of Respondents Who, as of 1975, Had Always Lived in the Same State

	(1) Unweighted		(2) Weighted	
	β	s	β	s
AGEH	0.012*	0.004	0.017*	0.005
EDUC	-0.009	0.040	-0.103*	0.050
MARR	-0.331*	0.160	-0.237	0.232
POOR	0.150*	0.110	0.327*	0.153
SEXH	-0.012	0.023	0.046	0.235
FOOD	-0.035*	0.021	-0.076	0.074
RELG	-0.003	0.030	-0.501*	0.286
CHEM	0.249	0.247	-0.147	0.332
TSPM5	0.4×10^{-4}	0.061	0.002	0.003
TSPM4	0.001	0.033	0.001	0.005
TSPM3	0.001	0.003	-0.001	0.004
TSPM2	0.003	0.016	-0.003	0.004
TSPM1	0.008*	0.004	-0.003	0.005
TSPM0	0.007	0.011	0.004	0.005
TSPM9	0.006	0.006	0.002	0.005
Constant	0.444		-0.690	
R^2	0.184		0.129	
S.E.	1.032		0.969	
F	(12,347) = 6.481		(15,210) = 4.082	

*Statistically significant at the 0.05 level of the one-tailed t-test.

unknown. The heteroskedasticity appears to be by far the most prominent for those estimated expressions having coefficients of determination less than 0.10.

It is widely thought that pollution-induced chronic illness is usually the result of cumulative, rather than instantaneous, exposures. Previously we have taken the position that, if only non-movers are represented in the sample, air pollution exposures during the year for which the respondent reports his behavior and status serve as adequate proxies for differences among respondents in cumulative exposures. If this position is at all tenuous, we have available the data to remedy it at least partially; that is, we have available information on respondent residential locations and air pollution exposures for a number of years. Table 8 presents some preliminary results involving an attempt to estimate the lagged effects of total suspended particulates upon the duration of chronic illness for 1975 respondents who have always lived in the same state. Since it is unclear exactly what a lagged effect of pollution upon the duration of an illness means, we exploit the high intercorrelation between LDSA and DSAB and interpret the expressions in terms of the lagged effects of air pollution upon the severity of chronic illness. As in earlier tables, the integers attached to the acronym for mean total suspended particulates refer to the year. Thus, for example, TSPMO refers to particulate concentrations in 1970.

The expressions presented in Table 8 have involved no tinkering: these are the first expressions having LDSA as a dependent variable that have used either of these samples. Expression (1) is an unweighted lag in which earlier air pollution concentrations are simply entered as additional explanatory variables. In spite of the very high simple correlation coefficients ($\gtrsim 0.80$) among the air pollution values of the various years, at least one year (1971) is statistically significant. Moreover, the magnitude of the coefficient increases from 1975 to 1971, and then starts to decline. We have no explanation for this rather neat pattern and tend to suspect that its very neatness in an anamoly that would fail to emerge in expressions estimated from other samples. These other samples have not yet been exploited.

The air pollution series in expression (2) has more structure imposed on it. In particular, the series is assumed to follow a geometric lag distribution where the coefficients decline in fixed proportions, causing the impact of more distant air pollution concentrations to become progressively smaller. Clearly, expression (2) does not accord any importance to total suspended particulates. However, this does not mean that all weighted lag structures will give similar results. Estimation techniques are available that allow one to fit polynomial structures of any degree. These techniques have not been applied here.

In concluding these remarks about dose-response functions, we must make explicit a feature of the SRC data set that could readily cause the morbidity effects of air pollution and other negative health influences to be biased downward. This possible bias is due to the retrospective feature of the SRC data: living individuals are questioned about their behavior and status during the preceding year. The problem arises because some potential respondents who were alive during the preceding year are dead by the time the interview occurs. Presumedly, those who died would tend to be those who were most seriously ill. If air pollution and other negative health influences contribute to this seriousness, or if those who are most seriously ill are most sensitive to air pollution, then the health impact of air pollution will be understated. Thus, the dose-response functions presented here are relevant only for those individuals who managed to survive over the time interval which the interviews described and the calendar date at which the interviews occurred. This qualification applies to all sections of this report where the SRC data is exploited. It is not a minor qualification since approximately five percent of the respondents died between interview years.

5.4 Recursive Estimates of the Effect of Air Pollution Upon Health, Labor Earnings, and Hours of Work

In the past decade, a number of empirical studies have appeared that describe the effect of health status upon labor productivity, where productivity effects are measured in lost earnings and work-time.[12] At the same time, numerous epidemiological studies that attempt to associate health status with air pollution have been published.[13] Thus far, no one has tried to combine the two study objectives in order to grasp the effect of air pollution upon either of the aforementioned measures of labor productivity. This section is a first attempt to do so. Labor productivity effects have never been explicitly included in quantifications of the benefits of national air pollution control efforts. Our results suggest that these productivity effects could constitute a significant portion of these benefits and are certainly worthy of further study.

In spite of a number of limitations which will later be exposed, the section has at least three unusual, if not utterly novel, features. First, although it treats health status as an exogenous rather than endogenous variable, a structural equation for health status is specified. This contrasts with nearly all epdemiological studies, where the analysis is confined to reduced-form health status, making any direct assignment of health effects to air pollution an extremely tenuous operation. Second, the health parameters in this section are estimated in the context of structural expressions for hourly earnings and annual hours of work. Finally, possible differences in effects of air pollution upon crude measures of acute and chronic generalized health status are recognized. The null hypothesis is that air pollution, by enhancing initial susceptibility and by making recovery more difficult, causes acute and/or chronic health problems. This, of course, was the theme of the previous section. In this section, we wish to ascertain the impact, if any, of these air pollution-induced health adversities upon earnings rates and hours worked. Thus through the intermediary of any health problems it induces, air pollution can be said to indluence labor productivity.

Even though health is treated as being exogenously determined, the Grossman (1972) model of health production can serve as the analytical foundation of the expressions to be estimated.[14] This model views the individual as the producer, via his selections of mixes of market-purchased goods and his own time, of health status. Within the context of this approach, earnings

Table 5.9

Simple Correlation Coefficients Between Labor Supply and Certain Other Variables for a 1970 Sample

	WAGE	WORK	BDALO	LTWK	ICTR	UION	RINC
WORK	0.085	1.000	0.287	0.123	-0.629	0.101	0.479
BDALO	0.235		1.000		-0.167		0.656
LTWK	-0.038		0.070		0.155		0.054
UION	-0.039		0.108		-0.131		0.070
RINC	0.465		0.656		-0.268		1.000
ICTR	-0.421	-0.629	-0.167		1.000		-0.268
ACUT	-0.042	0.012	-0.170		-0.102		-0.079
DSAB	-0.134	-0.468	-0.156		0.325		-0.227
LDSA	-0.141	-0.441	-0.143		0.303		-0.190
AGEH	0.017	-0.174	0.197		0.156		0.107
CITY	0.094	0.016	0.018		-0.057		0.092
EDUC	0.165	0.323	0.493		-0.172		0.465
FEDU	0.044	0.131	0.244		-0.132		0.148
FMSZ	-0.038	-0.046	-0.424		0.165		0.094
POOR	-0.114	-0.116	-0.126		0.094		-0.058
RACE	-0.014	0.008	0.139		-0.199		0.018
SEXH	0.072	0.448	0.360		-0.284		0.480
EXER	0.072	0.161	0.193		-0.150		0.217
FOOD	-0.007	-0.078	-0.058		-0.239		-0.061
RISK	0.139	0.235	0.538		-0.138		0.427
INSR	0.088	0.505	0.439		-0.413		0.440
CHEM	-0.012	-0.033	-0.082		0.040		-0.022
SULT	-0.037	-0.174	-0.163		0.085		-0.183
SULM	-0.038	-0.137	-0.113		0.083		-0.134
SULN	-0.056	-0.109	-0.077		0.048		-0.127
TSPT	0.002	0.081	0.087		0.066		0.075
TSPM	0.046	-0.005	0.086		0.044		0.130
TSPN	0.058	0.009	0.122		0.046		0.170

rates depend on various forms of investment in human "capital" (e.g., education, prior lifestyles, and medical inputs) and labor market conditions; and the time supplied to the labor market depends on the individual's hourly earnings and the quantities of goods and time desired for household production and consumption. Health states depend on the prior resources the individual has devoted to their production.

Except for certain of the environmental variables, the data used to estimate the model consist of four distinct samples drawn from the 1969, 1970 and 1971 SRC interview data. Several variables, defined in Table 5.1, are used in this section that were not used in the preceding section. For one of the samples, Table 5.9 provides the simple correlation coefficients between these additional variables and some of the other previously used variables. Representative means and standard deviations for the additional variables are available in Table 5.2.

Table 5.9 gives little attention to LTWK and UION because our major interest in them is their association with WORK, WAGE, and RINC. Absenteeism was checked in this sample but apparently none of the respondents would admit to being absent from work for reasons other than sickness. As was noted in Table 5.4, where 81.1 percent of the respondents had annual asset incomes of no more than $500, most of the respondents' annual incomes not earned during the current year appear to be governmental transfer payments. This accounts for the negative and high correlations between ICTR and RINC and WAGE. Note also in Table 5.9 that the simple correlations between the two chronic illness measures, DSAB and LDSA, and WORK and RINC are quite high.

The household head's annual number of work hours, WORK, and his hourly earnings, WAGE, are used as the empirical representations of the endogenous variables in the model. Remember from the definitions of Table 5.1 that WAGE is an approximation of the marginal, rather than the average, wage rate. Annual number of work hours is used as the sole measure of labor supply because the sample contains no information on the seasonal distribution of hours when working. Neither vacation time nor sick time is included in annual hours worked, even if the individual was paid during these times.

The system to be estimated for each sample consists of four expressions: a chronic illness expression; an acute illness expression; a wage expression; and a labor supply expression. A representation, in implicit form, of this structural system is as follows:

1. LDSA = f(Biological and social endowments, Lifestyles, Medical care, Environmental). (5.1)

2. ACUT = g(LDSA, Biological and social endowments, Lifestyles, Medical care, Environmental). (5.2)

3. WAGE = h(LDSA, ACUT, [illegible] of Living, Experience, Biological and social endc[illegible] (5.3)

4. WORK = k(WAGE, LDSA, ACUT, Transfer income, Wealth). (5.4)

As structured, this system is obviously recursive.

A great deal of research is available [e.g., Lazear (1976)] showing that earnings are positively related to formal and informal schooling. Good health is here viewed as having effects on earnings analogous to the effects of increased schooling; that is, good health increases the individual's marginal value productivity and therefore raises his marginal earnings. In addition, previous good health may have had an indirect effect on earnings by easing the task of achieving schooling success and thereby ultimately improving the individual's productivity and associated earnings. The EDUC and LOCC variables in (5.3) are intended to capture the effects of training upon earnings. They may also reflect, in part, the influence of past health status. The health status variables, ACUT, DSAB, and LDSA, in (5.3) register the effect of current health status, via the effect on productivity, upon earnings. Since chronic illnesses reflect long duration, as opposed to temporary, reductions in productivity, we expect wages to exhibit greater responsiveness to the chronic illness variables than to the acute illness variable.

In addition to the aforementioned variables, the marginal earnings expression includes variables representing the 1970 cost-of living in the county of residence as well as variables representing the individual's race and sex. If, as is frequently asserted, being non-white or female negatively influences marginal earnings, either labor market discrimination or less market productivity in the current period could account for the influence.[15] The structural system we employ is incapable of distinguishing between the two possible influences.

Cost-of-living, BDAL, in the county of residence is accounted for in (5.3) because it is real marginal earnings, rather than money earnings, that limit the extent to which the individual is able to satisfy his cravings and yearnings.

As Mincer (1970) and others have shown, earnings expressions similar to (5.3) should be semi-logarithmic, where the dependent variable is the logarithm of the earnings term. In this paper, we presume the earnings expression to be linear in the original variables. This presumption was adopted in order to obtain a sample of individuals possessing reasonable variability in the values of the health variables, earnings, and hours worked. If, in order to avoid having to assign positive earnings to individuals who really had zero earnings, only individuals who actually had positive earnings were included in the sample, the variability of the chronic disability measures would have been substantially reduced, thus requiring that inferences about the influence of air pollution on health, earnings, and hours worked be drawn from the relatively few remaining individuals whose health status and work patterns differed substantially from the mean. Moreover, dropping individuals with zero earnings from the sample would have meant that those individuals with long-standing and/or severe chronic health problems would be excluded.

Expression (5.4) the annual hours worked or labor supply expression, is consistent with the treatments of health capital in Grossman (1972). Improvements in health states increase the total time available for work and for consumption. With real earnings and consumption opportunities held constant, the consumer would be inefficient, assuming he was initially in equilibrium, if he allocated all this additional time solely to consumption. This is because the ratio of consumption time to work would rise, causing the marginal value of consumption time to become less than the marginal earnings that could be obtained. To recover equilibrium the individual would have to devote the additional time to both work and consumption. We therefore expect the amount of work time to increase with improvements in health status.

In addition, since health status is assumed to be exogenous, an improvement in health would increase the wage rate as well as the pecuniary equivalent of time spent in consumption. In terms of the household production approach to consumer theory, "full income" would be increased. The health improvement therefore would constitute a pure income effect, causing the individual to increase the value he attaches to any unit of consumption time. This increase in the value of consumption time would cause the individual to increase his demand for those marketed goods permitting him to use this more highly valued consumption time with greater effectiveness. The purchase of these marketed goods requires that he obtain more income, and therefore that he increase his work time.

An increase in income not earned in the current period, ICTR, would also result in a pure income effect. However, because the additional income is not a consequence of improvements in work productivity, the value of work time relative to consumption time decreases, assuming the wage rate and health status to be invariant. The result is that with an increase in income not earned in the current period, the individual must reduce work time in order to restore equilibrium.

The preceding remarks indicate why the sign of the marginal hourly earnings variable, WAGE, in (5.4) is ambiguous. An increase in marginal hourly earnings would increase the value of work time relative to the value of consumption time, causing the former type of time to be substituted for the latter. However, the increase in marginal hourly earnings has simultaneously increased the individual's "full income," causing the value he attaches to any given unit of consumption time to increase. Whether the increase in the value of consumption time exceeds the increase in the value of work time is an empirical question.

Since the immediately preceding remarks refer only to real marginal hourly earnings, (5.4) includes BDALO, the cost-of living index, in order to control differences in real earnings among counties of residence.

The four-equation system, in which acute and chronic illnesses are exogenously determined, represents a strictly recursive system. First, health status is determined independently of hourly earnings and hours worked, and then health status is used to determine hourly earnings and hours worked. Similarly, hourly earnings are determined independently of

Table 5.10a

Empirical Results for a 1971 Sample[a,b]
Recursive Labor Supply

1. $$\text{LDSA} = -1.018 + \underset{(0.04)}{0.06}(\text{NCIG}) - \underset{(0.07)}{0.17}(\text{DSAB})^{*} + \underset{(0.68)}{0.05}(\text{RELG}) - \underset{(0.19)}{0.40}(\text{CITY})^{*}$$

$$- \underset{(0.04)}{0.05}(\text{FEDU}) + \underset{(0.05)}{0.06}(\text{EDUC}) + \underset{(0.19)}{0.15}(\text{FOOD}) + \underset{(0.006)}{0.025}(\text{AGEH})^{*}$$

$$+ \underset{(0.002)}{0.001}(\text{COLD}) + \underset{(0.002)}{0.001}(\text{ULTV}) - \underset{(0.003)}{0.002}(\text{SULM}) + \underset{(0.002)}{0.005}(\text{NOXM})^{*}$$

$$+ \underset{(0.004)}{0.008}(\text{TSPM})^{*}$$

$R^2 = 0.21$; $F(13,\ 305) = 5.76$; S.E. = 1.56

2. $$\text{ACUT} = 35.397 + \underset{(6.20)}{0.25}(\text{NCIG}) - \underset{(9.40)}{0.38}(\text{DSAB}) - \underset{(7.59)}{4.98}(\text{LDSA}) - \underset{(35.76)}{47.39}(\text{EXER})$$

$$+ \underset{(4.85)}{5.14}(\text{EDUC}) - \underset{(0.87)}{0.36}(\text{AGEH}) + \underset{(0.27)}{0.06}(\text{COLD}) - \underset{(0.02)}{0.01}(\text{ULTV}) - \underset{(0.44)}{0.38}(\text{SULM})$$

$$+ \underset{(0.31)}{0.06}(\text{NOXM}) + \underset{(0.52)}{0.49}(\text{TSPM})$$

$R^2 = 0.15$; $F(11,\ 307) = 0.43$; S.E. = 215.15

3. $$\text{WAGE} = -5.619 + \underset{(0.06)}{0.08}(\text{BDALO}) + \underset{(0.10)}{0.02}(\text{ACUT}) - \underset{(44.41)}{156.70}(\text{SEXH})^{*} + \underset{(10.81)}{33.46}(\text{LOCC})^{*}$$

$$- \underset{(14.35)}{3.00}(\text{DSAB}) - \underset{(15.72)}{41.73}(\text{LDSA})^{*} + \underset{(10.20)}{20.31}(\text{EDUC})^{*} + \underset{(32.08)}{50.80}(\text{RACE})$$

$R^2 = 0.20$; $F(8,\ 310) = 9.72$; S.E. = 367.66

4. $$\text{WORK} = 2011.671 - \underset{(0.15)}{0.12}(\text{BDALO}) - \underset{(0.23)}{0.01}(\text{ACUT}) - \underset{(0.01)}{0.05}(\text{ICTR})^{*} + \underset{(26.35)}{123.40}(\text{SVGS})^{*}$$

$$- \underset{(34.26)}{10.55}(\text{DSAB}) - \underset{(36.52)}{212.00}(\text{LDSA})^{*} + \underset{(0.13)}{0.50}(\text{WAGE})^{*}$$

$R^2 = 0.35$; $F(7,\ 311) = 23.68$; S.E. = 865.51

(continued)

Table 5.10a
(continued)

[a] Standard errors are in parentheses.

[b] The sample includes only respondents who resided in 50 large U.S. cities.

*Significant at the 0.05 level of the one-tailed t-test.

Table 5.10b

Empirical Results for a 1970 Sample[a]
Recursive Labor Supply

1.

$$\text{LDSA} = 2.980 + \underset{(0.035)}{0.554}(\text{DSAB})^{**} + \underset{0.004}{0.005}(\text{AGEH}) + \underset{0.029}{0.013}(\text{EDUC}) - \underset{0.037}{0.044}(\text{FEDU})$$

$$- \underset{0.103}{0.069}(\text{POOR}) + \underset{(0.488)}{0.072}(\text{RACE}) + \underset{(0.114)}{0.139}(\text{SEXH}) - \underset{(0.975)}{0.902}(\text{FOOD}) - \underset{(0.129)}{0.454}(\text{INSR})^{*}$$

$$- \underset{(0.575)}{1.645}(\text{CHEM})^{*} + \underset{(0.0011)}{0.0028}(\text{TSPM})^{*}$$

$R^2 = 0.525;\ F(11,388) = 38.920;\ \text{S.E.} = 0.964$

2.

$$\text{ACUT} = 165.208 + \underset{(13.34)}{39.52}(\text{LDSA})^{*} - \underset{(1.312)}{1.421}(\text{AGEH}) - \underset{(39.16)}{16.92}(\text{SEXH}) - \underset{(0.118)}{0.086}(\text{CIGE})^{b}$$

$$- \underset{(40.11)}{78.40}(\text{EXER})^{*} - \underset{(0.033)}{0.105}(\text{FOOD})^{*} - \underset{(13.26)}{38.84}(\text{RISK}) + \underset{(47.47)}{187.0}(\text{INSR})^{**} + \underset{(0.317)}{0.623}(\text{TSPM})^{*}$$

$R^2 = 0.195;\ F(10,389) = 6.139;\ \text{S.E.} = 204.462$

3.

$$\text{WAGE} = -132.318 - \underset{(14.440)}{25.930}(\text{LDSA})^{*} + \underset{(8.5780)}{24.070}(\text{EDUC})^{*} + \underset{(18.260)}{15.370}(\text{DSAB})$$

$$+ \underset{(6.079)}{26.880}(\text{FMSZ})^{*} + \underset{(6.138)}{42.380}(\text{BDALO})^{*} + \underset{(22.130)}{52.950}(\text{LOCC})^{*} - \underset{(33.88)}{7.163}(\text{LTWK})$$

$$+ \underset{(34.580)}{66.090}(\text{UION})^{*} + \underset{(34.22)}{47.60}(\text{RACE})$$

$R^2 = 0.408;\ F(11,\ 388) = 24.28;\ \text{S.E.} = 258.908$

4.

$$\text{WORK} = 1266.680 - \underset{(27.220)}{163.900}(\text{LDSA})^{*} + \underset{(0.130)}{0.354}(\text{WAGE})^{*} + 44.260(\text{FMSZ})^{*}$$

$$+ \underset{(80.27)}{519.800}(\text{SEXH})^{*} - \underset{(0.022)}{0.272}(\text{ICTR})^{*} + \underset{(15.200)}{23.060}(\text{BDALO}) - \underset{(0.031)}{0.074}(\text{ACUT})^{*}$$

$R^2 = 0.551;\ F(6,\ 393) = 80.41;\ \text{S.E.} = 663.196$

---

[a] Standard errors are in parentheses

[b] Annual family expenditures on cigarettes in dollars

*Significant at the 0.05 level of the one-tailed t-test.

**Significant at the 0.05 level of the two-tailed t-test.

Table 5.10c

Empirical Results for a 1971 Sample[a]
Recursive Labor Supply

1.

$$\text{LDSA} = 0.265 + \underset{(0.049)}{0.808}(\text{DSAB})^{**} + \underset{(0.004)}{0.007}(\text{AGEH})^{*} - \underset{(0.030)}{0.057}(\text{EDUC})^{*} - \underset{(0.035)}{0.044}(\text{FEDU})$$

$$+ \underset{(0.096)}{0.086}(\text{POOR}) - \underset{(0.119)}{0.057}(\text{RACE}) + \underset{(0.109)}{0.233}(\text{SEXH})^{**} - \underset{(0.81 \times 10^{-4})}{0.138\text{x}10^{-3}}(\text{FOOD})^{*}$$

$$- \underset{(0.125)}{0.496}(\text{INSR})^{*} - \underset{(0.916)}{0.002}(\text{CHEM}) + \underset{(0.0017)}{0.0019}(\text{TSPM})$$

$R^2 = 0.530$; $F(11,388) = 39.800$; S.E. = 0.904

2.

$$\text{ACUT} = 99.839 + \underset{(1.038)}{0.985}(\text{AGEH}) + \underset{(27.69)}{55.55}(\text{INSR})^{*} - \underset{(33.34)}{67.50}(\text{EXER})^{*} - 0.052(\text{FOOD})^{*}$$

$$- \underset{(11.29)}{10.59}(\text{RISK}) + \underset{(12.637)}{21.784}(\text{LDSA}) + \underset{(0.676)}{1.177}(\text{TSPM})^{*}$$

$R^2 = 0.091$; $F(10, 389) = 4.095$; S.E. = 236.224

3. Acute illness, ACUT, is assumed not to effect marginal hourly earnings. Marginal hourly earnings expressions that include ACUT as an explanatory variable have been estimated from three different samples. In each case, ACUT has proven to be statistically nonsignificant. See, for example, expression (3) of Table 5.10a.

4.

$$\text{WORK} = 682.263 - \underset{0.032}{0.078}(\text{ACUT})^{*} - \underset{(34.073)}{50.119}(\text{LDSA}) - \underset{(46.490)}{154.20}(\text{DSAB})^{*}$$

$$+ \underset{(14.02)}{30.460}(\text{FMSZ})^{*} + \underset{76.190}{515.000}(\text{SEXH})^{*} - \underset{(0.022)}{0.2771}(\text{ICTR})^{*} + \underset{7.148}{8.969}(\text{BDALO})$$

$R^2 = 0.562$; $F(7, 392) = 72.34$; S.E. = 654.473

---

[a] Standard errors are in parentheses.

*Significant at the 0.05 level of the one-tailed t-test.

**Significant at the 0.05 level of the two-tailed t-test.

Table 5.10d

Empirical Results for a 1969 Sample[a]
Recursive Labor Supply

1.
$$\text{LDSA} = -0.223 + \underset{(0.019)}{0.041}(\text{NCIG})^{*} - \underset{(0.104)}{0.090}(\text{INSR}) + \underset{(0.109)}{1.964}(\text{DSAB})^{*}$$

$$+ \underset{(0.078)}{0.1212}(\text{POOR}) - \underset{(0.199)}{0.098}(\text{EDUC}) + \underset{(0.52 \times 10^{-3})}{0.10 \times 10^{-4}}(\text{FOOD}) + \underset{(0.003)}{0.003}(\text{AGEH})$$

$$+ \underset{(0.0011)}{0.0013}(\text{TSPM}) + \underset{(0.0021)}{0.0018}(\text{SULM})$$

$R^2 = 0.478$; $F(9,390) = 39.69$; S.E. = 0.736

2.
$$\text{ACUT} = 447.874 + \underset{(35.56)}{16.61}(\text{MARR}) + \underset{(9.844)}{16.13}(\text{NCIG})^{*} - \underset{(47.09)}{88.71}(\text{INSR})^{*}$$

$$+ \underset{(16.08)}{47.04}(\text{LDSA})^{*} - \underset{(34.03)}{29.80}(\text{POOR}) - \underset{(0.231)}{0.564}(\text{FOOD})^{*} - \underset{(12.33)}{7.676}(\text{RISK})$$

$$- \underset{(1.456)}{1.306}(\text{AGEH}) - \underset{(0.706)}{0.963}(\text{TSPM}) + \underset{(0.925)}{1.518}(\text{SULM})^{*}$$

$R^2 = 0.095$; $F(10,389) = 3.139$; S.E. = 317.201

3.
$$\text{WAGE} = 49.305 + \underset{(2.869)}{1.275}(\text{FMSZ}) + \underset{(4.312)}{28.20}(\text{LOCC})^{*} - \underset{(7.203)}{12.07}(\text{LDSA})^{*}$$

$$+ \underset{(15.73)}{34.98}(\text{UION})^{*} - \underset{(38.13)}{24.16}(\text{EDUC}) + \underset{(16.95)}{136.6}(\text{RACE})^{*} + \underset{(17.75)}{116.9}(\text{SEXH})^{*}$$

$R^2 = 0.411$; $F(7,392) = 39.03$; S.E. = 143.265

4.
$$\text{WORK} = 1779.540 - \underset{(0.082)}{0.623}(\text{ACUT})^{*} + \underset{(10.15)}{25.87}(\text{FMSZ})^{*} - \underset{(0.026)}{0.077}(\text{ICTR})^{*}$$

$$+ \underset{(59.63)}{143.8}(\text{SVGS})^{*} - \underset{(25.90)}{15.02}(\text{LDSA}) - \underset{(0.165)}{0.277}(\text{WAGE})^{*} + \underset{(66.31)}{394.8}(\text{SEXH})^{*}$$

$R^2 = 0.253$; $F(7,392) = 18.95$; S.E. = 514.153

---

[a] Standard errors are in parentheses.

*Significant at the 0.05 level of the one-tailed t-test.

Table 5.11a

Labor Supply Effects of Air Pollution-Induced Chronic and/or Acute Illnesses

From Table 10a: Air Pollution Induced Chronic Illness Only

Effect of a One Unit Increase in Air Pollution Upon Labor Supply

	NOXM	TSPM
Direct Effect	-1.0600 hours	-1.6960 hours
Indirect (via WAGE) Effect	-0.1044 hours	-0.1669 hours
Total Effect	-1.1644 hours	-1.8629 hours

Sum of total effects from Table 10a expressions = -1.1644 - 1.8629 = -3.0273 hours.

From Table 10b: Air Pollution Induced Chronic and Acute Illnesses

Effect of a One Unit Increase in Air Pollution Upon Labor Supply via Direct Impact of Chronic Illness

	TSPM
Direct Effect	-0.458 hours
Indirect (via WAGE) Effect	-0.026 hours
Total Effect	-0.484 hours

Effect of a One Unit Increase in Air Pollution Upon Labor Supply via Impact of Chronic Illness on Acute Illness

	TSPM
Direct Effect	-0.017 hours
Indirect (via WAGE) Effect	Zero, by assumption
Total Effect	-0.017 hours

(continued)

Table 5.11a
(continued)

Effect of a One Unit Increase in Air Pollution Upon Labor Supply via Direct Impact of Acute Illness

	TSPM
Direct Effect	-0.046 hours
Indirect (via WAGE) Effect	Zero, by assumption
Total Effect	-0.046 hours

Sum of total effects from Table 10b expressions = -0.484 - 0.017 - 0.046 = 0.547 hours.

From Table 10c: Air Pollution Induced Acute Illness Only

Effect of a One Unit Increase in Air Pollution Upon Labor Supply via Direct Impact of Acute Illness

	TSPM
Direct Effect	-0.092 hours
Indirect (via WAGE) Effect	Zero, by assumption
Total Effect	-0.092 hours

From Table 10d: Air Pollution Induced Acute Illness Only

Effect of a One Unit Increase in Air Pollution Upon Labor Supply via Direct Impact of Acute Illness

	TSPM
Direct Effect	-0.9457 hours
Indirect (via WAGE) Effect	Zero, by assumption
Total Effect	-0.9457 hours

Table 5.11b

Value of Labor Supply Effects of Air Pollution-Induced Chronic and/or Acute Illnesses for Pollution Concentrations Two Standard Deviations Removed from the Mean Concentration

From Tables 10a and 11a

Mean air pollution ± two standard deviations

NOXM = 95.320 ± 82.470
TSPM = 115.818 ± 65.756

Labor supply effects

NOXM = (-1.164 hours)(± 82.470) ≡ 95.9951 hours
TSPM = (-1.8629 hours)(± 65.756) ≡ 122.4975 hours

Total Effects 218.4926 hours

Value of labor supply effects: ($2.92)(215) = $638.00

From Tables 10b and 11a

Mean air pollution ± two standard deviations

TSPM = 74.837 ± 87.864

Labor supply effects

TSPM = (-0.547 hours)(± 87.864) = 48.062 hours

Value of labor supply effects: ($3.23)(48) = $155.00

From Tables 10c and 11a

Mean air pollution ± two standard deviations

TSPM = 89.210 ± 55.938

Labor supply effects

TSPM = (-0.092 hours)(55.938) = 5.146 hours

Value of labor supply effects: ($3.59)(5.146) = $18.47

(continued)

Table 5.11b

Value of Labor Supply Effects of Air Pollution-Induced Chronic and/or Acute Illnesses for Pollution Concentrations Two Standard Deviations Removed from the Mean Concentration

From Tables 10d and 11a

Mean air pollution ± two standard deviations

SULM = 24.583 ± 46.690

Labor supply effects

SULM = (-0.9457 hours) (46.690) = 44.155 hours

Value of labor supply effects: ($3.32) (44.155) = $146.59

hours worked. Similarly, hourly earnings are determined independently of hours worked and then hours worked are determined from hourly earnings. As Kmenta (1971, p. 585) shows, estimation of a recursive system by ordinary least squares is equivalent to estimation by full information, maximum likelihood.

At this juncture, we wish to emphasize that the use of a single air pollution health effect, or effect of health on wages and/or hours worked, may be somewhat misleading. These effects may differ, for example, with age and numerous other variables. As one gets older, it may be that air pollution-induced health effects become progressively more severe, implying, for given levels of training and work experience, that the absolute effect of air pollution upon hourly earnings and hours worked increases with age. Ideally, this possibility makes it worthwhile to estimate separate expressions for different age groups. Otherwise, one obtains, as we do, a coefficient representing effects for neither old nor young people but simply a weighted average of the two from which it is impossible to disentangle the separate contributions of each group effect. In essence, in addition to all the other caveats that must be applied to the empirical results set forth here, one cannot blindly transfer these estimated air pollution-induced health, hourly earnings, and hours worked effects to other samples of individuals unless their age distribution is similar to the age distribution in these samples. If air pollution-induced effects also differ by other demographic attributes such as race and sex, a similar caution applies.

Tables 5.10a, 5.10b, 5.10c, and 5.10d present estimates of the chronic illness dose-response expressions, the acute illness dose-response expressions, the marginal earnings expressions, and the labor supply expressions. The samples of individuals used to estimate these expressions include housewives, retirees, and students, all of whom were assigned zero hours of acute illness by the Survey Research Center. These individuals constitute about twenty percent of the sample, thus imparting what is probably a substantial downward bias for these labor supply calculations in the estimated effects of air pollution upon acute illness. Failure to include these groups would have resulted, however, in the removal from the sample of a disproportionately high number of individuals with chronic illnesses.

Table 5.11a provides estimates of the direct and indirect effects upon labor supply, as measured by annual hours worked, of air pollution-induced acute and/or chronic illnesses. Assuming that the marginal hourly wage is an accurate representation of the market value of the worker's marginal productivity, these reduced work hours are valued in Table 5.11b at the marginal wage applying before the reduction in work hours. Apart from any issues dealing with the estimation procedures used to obtain each expression, the reader should be sensitive to the fact that assumptions stating that illness is unaffected by work-hours and/or wages underlie the calculations in these two sets of tables.

Detailed description of the calculation procedures in Tables 5.11a and 5.11b is both tedious and repetitious. In order to inform the reader of the procedure, we describe that applied to the material in Table 5.10a, leaving the reader the responsibiliy to invent for himself the procedures we have applied to Tables 5.10b, 5.10c, and 5.10d, which have resulted in the labor supply effect estimates set forth in Tables 5.11a and 5.11b.

Of the three air pollution variables in the chronic illness dose-response expression of Table 5.10a, two, NOXM and TSPM, have a positive sign and are statistically significant.[16] Making the already acknowledged dangerous assumption that each discrete interval of LDSA is slightly more than two years, or 830 days, the coefficient attached to NOXM implies that, on average, each unit increase in annual geometric mean concentrations of ambient nitrogen dioxide increases the length of chronic illness by 4.15 days.[17] Similarly, on average, each unit increase in annual geometric mean concentrations of ambient total suspended particulates increases the length of chronic illness by 6.64 days. Calculated at the arithmetic means, the elasticity of LDSA with respect to NOXM is 0.47, while the elasticity for TSPM is 0.95.

The signs of the coefficients for the non-health variables in the hourly earnings expression, (3), in Table 5.10a are in accord with *a priori* expectations. Except for BDALO and RACE, all are statistically significant at generally accepted levels. As for the health-related variables, neither acute illness nor the severity of disability appears to have an effect upon hourly earnings. However, the length of time over which the individual has been disabled has a substantial and statistically meaningful effect. An increase of two years in the length of time the individual suffers from a chronic illness reduces hourly earnings, on average, by 41.73 cents. When calculated at the means, the elasticity of WAGE with respect to LDSA is -0.17, implying that within the ranges of chronic illness time length and hourly earnings considered here, the response of hourly earnings to chronic illness is rather sluggish.

Using the above results for the effect of LDSA on WAGE, and the earlier results for the effect of NOXM and TSPM on LDSA, one can calculate the average effect of each of the two air pollutants upon hourly earnings. The 4.15 day effect of an additional unit of NOXM on LDSA is 0.50 percent of the 830 days said to constitute one unit of LDSA. Since a one unit increase in LDSA reduces hourly earnings by 41.73 cents, the average effect of an additional unit of NOXM on hourly earnings is (0.005)(-41.73) = -0.2087 cents. Performing the same calculations for TSPM, the average effect of an additional unit of total suspended particulates on hourly earnings is (0.008)(-41.73) = -0.3338 cents.

Among the non-health variables in the labor supply expression, (4) of Table 5.10a, only BDALO fails to be statistically significant. The coefficient for WAGE has a t-value slightly less than four, and it implies an elasticity of WORK with respect to WAGE of 0.12. This means that the substitution effect of a change in real earnings exceeds the income effect. The highly significant and negative coefficient attached to 1971 income secured by means other than 1971 labor, ICTR, is consistent with a substantial income effect that causes the individual to substitute consumption hours for work hours. The elasticity of WORK with respect to ICTR, when evaluated at the means of the variables, is -0.18.

The positive and statistically significant coefficient attached to WAGE implies that the length of time the individual has been chronically ill, LDSA, has an indirect as well as a direct effect upon the annual hours of work the individual supplies. This occurs because, as was observed in the references to (3) of Table 5.10a, LDSA lowers hourly earnings as well as having a powerful direct effect, according to (4), upon labor supply. Table 5.11a exhibits the direct, indirect, and total effects of NOXM and TSPM upon labor supply, as measured by annual hours worked. The total effect is an estimate of the coefficient for LDSA in a reduced form expression.

Assuming the average work day to be eight hours long, a one unit increase in LDSA directly brings about a 212 hour or 26.50 day reduction in annual working time. As earlier noted, 0.5 percent of a one unit change in LDSA is attributable to NOXM, while 0.8 percent of a similar change is due to TSPM. The direct effect of an additional unit of NOXM upon annual hours worked is therefore $(0.5 \times 10^{-2})(-212) = -1.06$ hours, while the direct effect of TSPM is $(0.8 \times 10^{-2})(-212) = -1.6960$ hours.

The indirect effect of air pollution upon labor supply is obtained by first recognizing that in (4) of Table 5.10a, each one cent change in hourly earnings generates an average change of the same sign of 0.50 in annual work hours. As was noted in the discussion of the empirical results for (3), an additional unit of TSPM reduces hourly earnings by 0.3338 cents. The indirect effect of an additional unit of TSPM upon annual work hours is then $(-0.3338)(0.50) = -0.1669$ hours; the indirect effect of an additional unit of NOXM on annual work hours is then $(-0.2087)(0.50) = -0.1044$.

On average, the total reduction in labor supply caused by a one unit increase in TSPM is 1.8629 hours, while the reduction for a one unit increase in NOXM is 1.1644 hours. Assuming the health of the representative individual in this sample to be exogenously determined, and that no potential interviewee died between the year for which behavior and status is recorded and the time of the interview, the total reduction in his annual hours worked caused by simultaneous one unit increases in NOXM and TSPM is then 3.0273 hours, i.e., approximately three hours. This last figure assumes that the effects of NOXM and TSPM are additive. Making the exceedingly strong assumptions that the effects of these two air pollutants upon hourly earnings and annual hours worked are constant over all ranges being considered and that the effect of hourly earnings upon annual hours worked is also constant, those individuals living in cities having air pollution concentrations two standard deviations removed from the mean concentration of the cities considered in this paper will have changes in annual hours worked of 95.9951 hours due to NOXM, and 122.4975 hours due to TSPM; that is, an individual who works and resides in an extremely clean city might work 218 hours more a year than the individual who works and resides in a city with average air pollution concentrations. Valuing these 218 hours at the marginal wage applying before the reduction in work hours, we have a loss in average total earnings of (218) (\$2.92) or \$638 per individual, a figure which, in spite of the grossness of our assumptions, is not in great discord with intuitive possibilities. Given our linearity assumption about the response of labor supply to air pollution, this results in \$1,276 in lost wages for an individual living in an extremely dirty location as compared to that same individual living in an extremely clean location.

In the preceding paragraphs, we have calculated:

$$\text{WAGE}\left(\frac{d\text{WORK}}{d\text{Pollution}}\right) = \frac{\partial\,\text{Illness}}{\partial\text{Pollution}} + \frac{\partial\,\text{WORK}}{\partial\text{WAGE}} \cdot \frac{\partial\,\text{WAGE}}{\partial\,\text{Ilness}} \cdot \frac{\partial\,\text{Illness}}{\partial\text{Pollution}}\ \text{WAGE}$$

As an alternative, we could readily have calculated:

$$\frac{d(\text{WORK}\cdot\ \text{WAGE})}{d\text{Pollution}} = \frac{d\text{WORK}}{d\text{Pollution}} \cdot \text{WAGE} + \frac{d\text{WAGE}}{d\text{Pollution}} \cdot \text{WORK}$$

This latter calculation procedure would yield results comparable to those obtained from the first calculation procedure. For example, the calculation for the expressions in Table 5.10a would have yielded (\$1,276)[-0.2087 + (-0.3338)] = \$692.

The lost wages occurring in the remaining three samples are considerably less. In the example from Table 5.10b, the difference between an extremely dirty location amounts to \$310, mainly because the total effect of air pollution upon chronic illness is much less in this sample (a coefficient of approximately 0.003 as opposed to a sum of coefficients of approximately 0.13) than in the sample of Table 5.10a. The lesser impact in the sample of Table 5.10b exists even though this sample includes a statistically significant acute illness effect of air pollution whereas the sample of Table 5.10a does not.

On the basis of the limited experience of these four samples, air pollution-induced acute illness appears to have a much smaller effect upon labor supply and productivity than does air pollution-induced chronic illness. This is reflected in the example from Table 5.10c as well as that from Table 5.10b. In the latter, although air pollution does significantly affect acute illness, its effect, via acute illness, upon labor supply is overwhelmed by the effect of air pollution-induced chronic illness. The sample of Table 5.10c must depend for its labor supply effects upon acute illness alone. Its magnitude is trivial relative to the air pollution induced chronic effects of Tables 5.10a and 5.10b. Note, however, that the money value of the labor supply effects of the air pollution-induced acute illness in Table 5.10d are nearly one-quarter of the total effects of the air pollution induced illnesses in Table 5.10a.

The empirical results set forth in this section suggest that air pollution, mainly via its influence on chronic illness, affects labor productivity, that at least the order of magnitude of the effect can be estimated within the immediate neighborhood of existing air pollution concentrations and health states, and that the estimates can be given meaning within a rigorous analytical framework. Nevertheless, the estimates we have obtained are basically reduced form estimates: the causally subsequent expressions relating to chronic and acute illnesses and marginal hourly earnings are simply substituted into the labor supply expression to obtain the total of the direct and indirect effects of air pollution induced health effects upon labor supply. This may be too extreme. We allow the individual's state-of-health to influence his earnings and his annual hours of work, but we do not permit these hours of work or earnings to influence his state-of-health.

Yet some empirical evidence exists that long hours of strenuous physical work may generate fatigue and thereby initiate or accentuate air pollution induced health effects.[18] Moreover, presumedly in order to try to capture socioeconomic and background influences for which they have no overt measures available, epidemiologists have often included earnings as an explanatory variable in dose-response functions. Even economists [e.g., Grossman (1972) and Cropper (1977)] have included wages or earnings in analytical statements of health production functions.

In a succeeding section, we attempt to establish empirically whether reciprocal relations exist between health states, work hours, and wages for a sample of respondents in the SRC data. Before doing so, however, we present an analytical model of consumer behavior which enables us to provide some a priori structure for these reciprocal relations. In particular, with this model we are able to interpret the estimated relations as demand functions for avoiding acute or chronic illnesses and predict the behavior of several of the function parameters. To the best of our knowledge, the model set forth in the next section is the first to conform to the common sense notion that health status is a direct source of utility as well as a factor that influences the efficiency of production and consumption activities.

5.5 A Model of the Effect of Air Pollution on the Demand for Health[19]

Let an individual obtain utility from two commodities: H, the discounted flow of health services in each period i, h_i; and Z, the present value of the stream of services per period of a composite commodity, z_i. Thus:

$$U = U(H,Z) \tag{5.5}$$

where

$$H = \sum_{i=0}^{I} \alpha_i h_i, \text{ and } Z = \sum_{i=0}^{I} \alpha_i z_i,$$

and α_i is the individual's discount factor for the ith period.

Presume that the individual has an initial health endowment, H_o, that was provided instantaneously in period 0. However, due to natural aging, this initial health stock depreciates exogenously over time as given by (5.6), where β_i is the proportion of H_o remaining in the ith period.

$$H_i = \beta_i H_o \tag{5.6}$$

The h_i and z_i are produced by linear homogeneous production functions f_j (j = H,Z) whose inputs are goods, X_{ij}, and time in each period i. Air pollution and other environmental goods are included among the X_{ij}. In general $\partial h_i/\partial X_i$, when i is pollution.

$$h_i = f_h(X_{hi},T_{hi}), \tag{5.7}$$

$$z_i = f_z(X_{zi}, T_{zi}), \tag{5.8}$$

where T_{hi} is the time allocated specifically to health care, and T_{zi} is leisure time. In the X_{hi}, no distinction is made between ameliorative and preventive medical care, since, if the ameliorative care returns the individual to his former health status, he is dropped back into the same risk pool he was in before receiving the ameliorative care.

We make a distinction between the time-based wage rate and an incentive payment based on the flow of productive services the individual provides. The latter is viewed as a supplement to the time-based salary. It is a reward varying directly with the effort the individual expends over and above that minimum expenditure necessary for him to keep his job. This distinction between time-based salary and incentive payments for non-prescribed effort expenditures allows us to discriminate between acute and chronic health effects insofar as they influence the efficiency of production and consumption activities. Acute health effects do not alter total earnings except when they reduce time on the job, whereas chronic effects alter both time on the job and total earnings for any given amount of time on the job.

Total incentive payments, M_i, are given by (5.9), where $g(\cdot)$ is a twice-differentiable, decreasing returns-to-scale production function, P is the incentive payment, and E and e are respectively stock and flow non-health environmental variables (e.g., schooling, services of a mate, air pollution that directly affects productivity, rather than via health, coffee, air conditioning, etc.) that may influence the ability to put forth effort. The c's are their respective unit prices. Note that M_i varies directly with the amount of output the individual's efforts produce, rather than the amount of effort he expends.

$$M_i = Pg(h_i, E_i, e_i, T_{Di}) - c_E E_i - c_e e_i. \tag{5.9}$$

In (5.9) T_{Di} represents time expended on other work activities in the the period, including household production. These activities are presumed to dissipate energies that could otherwise be devoted to work. Alternatively, one could include T_{Wi}, work time, rather than T_{Di} in (5.9) on the presumption that, beyond some positive time expenditure, additional work time causes fatigue and/or ennui.[20]

The individual's ith period time constraint is given by (5.10) where θ_i is Becker's (1967) "full-time," and T_{Wi} is work time.

$$\theta_i = T_{hi} + T_{zi} + T_{Di} + T_{Wi} \tag{5.10}$$

If p_h, p_z are the price indices of the goods used in the production of h and z, and if x_h, x_z are the average (= marginal) composite purchased good coefficients of h_i and z_i, then the individual's budget constraint over his planning horizon can be represented as:

$$\sum_{i=0}^{I} Y_i = T_{Wi}W + M_i - p_h x_h h_i - p_z x_z z_i = 0, \tag{5.11}$$

where Y_i is the ith period flow of non-earnings income, and W is the time-based wage rate.

Upon combining (5.10) and (5.11), assuming W represents the shadow-price of time, one obtains the "full" intertemporal wealth constraint, (5.12):

$$S = \sum_{i=0}^{I} \alpha_i[\theta_i W + Y_i + P_g(h_i, E_i, e_i, T_{D_i}) - (P_h X_h + WT_h)h_i$$

$$- (P_z X_z + WT_z)z_i - c_E E_i - c_e e_i] = 0 \qquad (5.12)$$

The optimal levels of H and Z, the optimal uses of stock and flow non-health environmental variables, and the utility-maximizing time allocations in each period are obtained by maximizing (5.5) subject to (5.12) with non-negativity constraints on H_o, Z, E, e, and T_D. There are thus 3I + 2 first-order conditions including the full-wealth constraint.

$$\sum_{i=0}^{I} \alpha_i U_H + \lambda[\sum_{i=0}^{I} \alpha_i(\beta_i P_{g_{h_i}} - \beta_i[P_h X_h + WT_h])] \leq 0; \qquad (5.13)$$

$$U_z + \lambda[-\Sigma \alpha_i(P_z X_z + WT_z)] \leq 0; \qquad (5.14)$$

$$P_{g_{T_{D_i}}} - W \leq 0; \qquad (5.15)$$

$$P_{g_{E_i}} - c_E \leq 0; \qquad (5.16)$$

$$P_{g_{e_i}} - c_e \leq 0; \qquad (5.17)$$

Assuming internal solutions, expression (5.13) can be rewritten to form (5.18):

$$\sum_{i=0}^{I} \alpha_i \beta_i (P_h X_h + WT_h) - \sum_{i=0}^{I} \alpha_i \beta_i P g_{h_i} = \sum_{i=0}^{I} \frac{\alpha_i U_H,}{\lambda} \qquad (5.18)$$

which says that the optimal state-of-health occurs where the present value of health is less than the capitalized cost by the value of the marginal utility of the health stock. Thus, the net price of health as an unput into the work process is the horizon period consumption price, $\Sigma \alpha_i \beta_i (P_h X_h + WT_h)$, less the pecuniary equivalent of marginal utility.

Upon combining (5.13) and (5.14), one obtains:

$$\sum_{i=0}^{I} \alpha_i \left(\frac{U_H}{U_Z}\right) = \frac{\sum_{i=0}^{I} \alpha_i \beta_i \, (P_h X_h + WT_h) - \sum_{i=0}^{I} \alpha_i \beta_i P_{g_{h_i}}}{- \sum_{i=0}^{I} \alpha_i (P_z X_z + WT_z)}, \tag{5.19}$$

$$\equiv c_h / c_z$$

which states that the marginal value product of health in work offsets the predetermined consumption price component. Thus, one consequence of the dual role of health is that, even though the time-based wage rate is fixed and the household production functions are linear homogeneous, the full shadow price of health in production or consumption is endogenous, dependent on the state-of-health demanded since the marginal product of better health, Pg_{h_i}, and the marginal utility of better health, U_H/λ, will decline as H_o increases.

To ascertain the response of health states demanded to changes in the parameters specified in the model and to formulate a demand function for health states, the first-order conditions (5.13) - (5.17) must be totally differentiated and the relevant partials for H_o calculated.

The response of health demand to own predetermined price, $\Sigma\alpha_i\beta_i(P_hX_h + WT_h)$ can be decomposed into compensated substitution and (full) income effects:

$$\frac{\partial h}{\partial[\Sigma \alpha_i \beta_i (P_h X_h + WT_h)]} = \frac{\partial \bar{h}}{\partial[\Sigma \alpha_i \beta_i (P_h X_h + WT_h)]} + h\left(\frac{\partial h}{\partial s}\right), \tag{5.20}$$

Since the first term on the right-hand side of (5.20) corresponds to a compensated price effect, the second-order conditions require it to be negative. It is unclear what the sign of $h\left(\frac{\partial h}{\partial s}\right)$ should be.

Under the assumption that the individual's price of time is equal to the time-based wage rate, the uncompensated substitution elasticity of health with respect to the time-based wage rate is:

$$G_{H_W} = \varepsilon^{*}_{H} c_H (\lambda H - \lambda_z) + \sum_{i=0}^{I} \phi_{H_W} + \left[\sum_{i=0}^{I} \alpha_i T_{D_i} W \varepsilon_s \right] / s \tag{5.21}$$

where $\varepsilon^{*}_{H} c_H$ is the own compensated price elasticity of health; ε_s is the (full) income elasticity of health stock, H; and the consumption price time intensities are defined as:

$$\gamma_H \equiv \left(\sum_{i=0}^{I} \alpha_i WT_H \right) / c_H \tag{5.22}$$

$$\gamma_z \equiv (\Sigma\, \alpha_i WT_z)/(c_z = p_z x_z + WT_z) \tag{5.23}$$

In (5.21) a compensated increase in the time-based wage rate reduces the demand for the absence of acute health effects (causing the value of freedom from air pollution exposures to be reduced) if the individual's production of freedom from acute health effects is more time-intensive than is his production of other goods from which he obtains utility. This is because the second-order conditions require that $\varepsilon^{*}_{H}c_{H} < 0$. Although we can only speculate, activities such as daily exercise programs and the careful preparation of healthy menus do seem more time-intensive than reading a novel or eating at the local fast-food emporium. Even if the consumption price time intensities are equal, i.e., $\gamma_H = \gamma_Z$, an increase in W might still reduce the demand for freedom from acute health effects, since, from (5.22) and (5.23), $\gamma_H > \gamma_Z$ as $c_H > c_Z$.

The second-order conditions imply that there will exist a discrepancy between the observed income elasticity of health status and the "true" income elasticity. In fact, the former is likely to be less than the latter because the data used to calculate the observed elasticity will frequently be unable to distinguish between the time-based and the incentive payment components of the total wage. These two components imply that the individual's budget constraint will be nonlinear since chronic health status influences the ability of the individual to provide those productive services rewarded by incentive payments. This downward bias further implies that estimates of the demand for the absence of air pollution induced chronic health effects will also be biased downward whenever the data do not allow a distinction between the two earnings components. If an exogenous reduction in air pollution increases the optimal degree of absence of chronic illness, the marginal productivity portion of the full shadow price of health diminishes, assuming that the supply of effort is negligibly reduced by the additional earnings. The shadow price of the health stock therefore rises.

A second general consequence of the contribution of freedom from chronic illness to incentive payments is that changes in education and similar factors related to the provision of productive services will influence the shadow price of health by altering horizon period productivity. In turn, these factors will affect the value the individual attaches to the absence of air pollution-induced chronic illness. In short, the individual's demand for freedom from air pollution exposures will be related to his education, job experience, and other influences on his productivity.

The uncompensated elasticity of freedom from chronic health effects with respect to the price, c_j, of any of the aforementioned factors related to the provision of productive services is ambiguous, however. This elasticity is given by (5.24), where q_j refers to one of these productive services.

$$\varepsilon_{H_{c_j}} = \sum_{i=0}^{I} \phi_{H_{c_j}} + \frac{\sum_{i=0}^{I} \alpha_j q_j c_j \varepsilon_s)}{s} \tag{5.24}$$

The sign of this expression depends on whether the factor in question is a substitute for ($\phi > 0$, as with education), or a complement of ($\phi < 0$ as perhaps with comfortable job surroundings) freedom from chronic illness. For example, assuming non-inferiority ($\varepsilon_s > 0$), if freedom from chronic illness and the services of a mate are (imperfect) substitutes in the provision of productive services, then a compensated increase in c_j would raise the demand for health; the sign of the uncompensated effect, however, would depend on the magnitude of ε_s and the share of the costs of the services of a mate in full income.

The effect of a change in the price, P, of the output of productive services is also ambiguous. Expressed in elasticity terms, this effect is:

$$\varepsilon_{Hp} = -\varepsilon^*_H c_H \sum_{i=0}^{I} \alpha_i \beta_i P g_i / c_H + \sum_j \sum_{i=0} \phi H_{c_j} + \sum_{i=0}^{I} P g_i / R \varepsilon_R .$$

While an increase in P raises the marginal value product of good health, thus lowering c_H, and increases incentive-based income, the value of the output contributions of the other input factors in g(•) also rises. The sign of the compensated substitution effect will thus depend on the complementarity-substitution relations between freedom from chronic health effects and other inputs.

Accounting for the preceding development, we can express the demand for freedom from chronic and acute illnesses in terms of two functions differing according to whether we are considering acute or chronic illness. Both of these functions will involve arguments, however, relating to predetermined variables that influence the price of time, in addition to variables that relate to production and consumption activities. Thus, for the willingness to accept chronic illness, we can write the demand function as:

$H_{LDSA} = \mu_1$ (Time-based wage, Incentive payment, Non-earnings income, Environmental variables, Cost-of-living, Endowment variables).

In the case of the demand for acute illness, the demand function, $\mu_2(\cdot)$, for H_{ACUT} is similar to H_{LDSA} above, except that the term for incentive payments is deleted.

5.6 Some Empirical Results: The Demand for Freedom from Air Pollution-Induced Acute and Chronic Illness

The model of the preceding section implies that changes in the willingness to accept acute illness will result in changes in work time alone, although the extent of the change will depend on other parameters such as the time-based wage rate, transfer income, and assorted background variables.

In contrast, the wage rate is endogenous in the demand for freedom from chronic illness, since the extent of chronic illness determines, in part, the wage rate. Thus, although the wage rate is determined outside the system in the demand for freedom from acute illness, it is determined within the system in the demand for freedom from chronic illness. This implies that we can treat the demand to avoid acute illness as a single expression, but must account for the simultaneity between the wage rage and chronic illness when estimating the demand to avoid chronic illness.[21] In the latter case, we must resort to simultaneous equation estimation procedures. Here we adopt two-stage least squares.[22]

The appropriate expressions to calculate the pecuniary amounts the individual would have to receive to be willing to pay to avoid an increase in acute or chronic illness are, respectively, (5.21) and (5.24) of the previous section. Calculation of these expressions is clearly rather complex. As an alternative, we have calculated this willingness to accept as:

$$\frac{d(\text{Income})}{d(\text{Pollution})} = \frac{d(\text{Illness time})}{d(\text{Pollution})}\,(\text{Price of time})(\text{Illness time})$$

Upon reflection, this proposed method of calculation seems no different than the procedure employed to calculate the pecuniary equivalent of the recursive effects of air pollution upon labor supply. A somewhat subtle difference does nonetheless exist. In particular, a difference exists in the definition of illness time and its response to pollution variation: The recursive estimates dealt only with the physical effects of air pollution, while illness time in the above expression represents the individual's utility-maximizing illness time. When estimating dose-response expressions, we included as explanatory variables only predetermined variables either that described the individual's current health status or were *a priori* physical determinants of changes in this status. In contrast, when estimating the individual's demand expression for willingness to pay to avoid illness, we include variables such as the time-based wage rate, transfer payments, incentive payments, etc., that influence the sacrifices the individual is willing to make in order to avoid illness time. For consistency, and only when we have no alternative, we even sometimes reinterpret the meanings of identical explanatory variables that appear in both the dose-response expressions and the demand expressions. Thus, INSR, which was conveniently interpreted as a proxy for the availability of medical care in the dose-response expressions, will be interpreted in the demand expressions as a proxy for the price that the individual faces for a given quantity of medical care.

In the analytical model of the preceding section, increased air pollution reduces the flow of health services, and, as a consequence, reduces utility and usually increases the marginal product of particular health investments. These effects are opposing, causing the sign to be expected for the coefficients attached to the pollution variables to be ambiguous. However, pollution also causes the cost of supplying a given health status to increase. The result is that the income the individual is willing to forego to avoid pollution-induced illness is positive. We therefore expect the signs attached to the pollution coefficients in the demand expressions to be unambiguously positive.

Table 5.12 below presents three estimated demand expressions relating to acute illness for two different samples drawn from the 1971 SRC data. These samples include housewives, retirees and students, all of whom were assigned zero hours of acute illness by the SRC. The expressions are linear in the original variables. Expressions (1) and (3) were estimated from the same example. Only in the first two expressions is at least one of the air pollution variables statistically significant. The individual's time-based wage, which was measured as his hourly earnings on his regular job, appears to have no influence on his demand for avoiding increased hours of acute illness. Neither does annual work hours nor cigarette expenditures. As previously noted, substantial measurement error is involved in CIGE. People who participate in energetic activities and have adequate diets tend to have greater demands for the avoidance of acute illness, as do those who are risk averse. Older people and those who face lower prices for medical care seem less willing to pay to avoid additional acute illness. As in the dose-response expressions for acute illness, the sign attached to INSR is puzzling. Additional income, the acquisition of which does not involve any current time, increases the demand for acute illness avoidance.

In expressions (1) and (2), each additional unit of TSPM results, respecively, in an additional 1.212 and 0.796 additional optimal acute annual hours of illness. In expression (1) of Table 5.12, the arithmetic mean of WAGE is \$3.62, meaning that the representative individual would, on average, be willing to pay an additional \$4.39 annually to avoid one additional unit of TSPM. The arithmetic mean of WAGE for expression (2) in Table 5.12 is \$3.58. Thus, the representative individual in this sample would be indifferent between paying \$2.85 and an additional unit of TSPM. In expression (2), the arithmetic mean TSPM is 87.315 and 54.749 units of TSPM is two standard deviations removed from this mean. Assuming a linear extrapolation of the preceding marginal (= average) willingness to pay of \$2.85 for avoiding an additional unit of TSPM to be valid, the representative individual in expression (2) would be willing to pay \$312 annually to avoid the additional hours of acute illness associated with living in a location where TSPM is extremely high as opposed to being extremely low. A similar calculation for the representative individual in expression (1) indicates that he would be willing to pay \$457.97 in 1971 dollars annually in order to avoid a similar fate.

The basic calculations from the willingness to pay to avoid chronic illness expressions in Tables 5.13a and 5.13b are identical to the procedures used for the willingness to pay to avoid acute illness expressions of Table 5.12. The sole difference is the use of the arithmetic mean value $\widehat{WAGE}$ rather than WAGE. Table 5.13a holds no special surprises except for the sign attached to the statistically significant coefficients of DSAB. Nonetheless, the sign is consistent with a finding of Hamushek and Quigley (1978) that disabliities appear to affect negatively the earnings of the blue-collar workers but have little, if any, effect on the earnings of (presumedly) higher paid white-collar workers.

The estimates in Table 5.13b indicate a reduced quantity demanded of chronic illness avoidance with an increase in age, and an increased quantity demanded with reduced prices for medical care. The significance of the

Table 5.12

Willingness to Pay to Avoid Acute Illness

Sample	(1)		(2)		(3)	
Variable	β 1971	s	β 1971	s	β 1971	s
WORK	0.007	0.018	-0.007	0.010	0.012	0.018
WAGE	0.047	0.032	-0.016	0.020	0.039	0.052
CIGE	-0.057	0.094	-0.108	0.118	-0.067	0.095
EXER	-66.990*	33.320	-30.033	40.019	-60.520*	33.620
FOOD	-0.052*	0.024	-0.115*	0.033	-0.052*	0.024
RISK	-10.460	11.250	-40.020*	13.420	-12.680	11.360
AGEH	0.955	1.034	-0.506	1.286	1.246*	0.742
INSR	54.85**	27.58	161.800**	47.230	63.480*	27.890
ICTR	-0.244*	0.022	-0.278*	0.022	-0.233*	0.021
TSPM	1.212*	0.668	0.796*	0.384	0.500	0.478
SULM	-0.610	0.419				
Constant	99.057		182.339		128.082	
R^2	0.091		0.086		0.089	
S.E.	245.647		267.306		258.336	
F	(10,391) = 3.094		(9,390) = 4.112		(9,390) = 3.475	

*Statistically significant at the 0.05 level of the one-tailed t-test.

**Statistically significant at the 0.05 level of the two-tailed t-test.

Table 5.13a

Two-Stage Least Squares Estimates of WAGE Expressions for Chronic Illness

Sample	(1)		(2)	
	1970		1971	
Variable	β	s	β	s
EDUC	31.730*	8.735	23.740*	5.307
WORK	0.016	0.021	0.0017*	0.0011
DSAB	179.200**	50.230	35.670**	17.220
FMSZ	33.610*	6.293	63.790*	37.980
BDALO	40.470*	6.075	14.610*	3.447
HMPN[b]	0.554**	0.218	0.642	0.927
LTWK	-17.430	33.520	-35.160	32.070
ABSN[a]	-5.401	44.010	-	-
UION	87.320*	34.620	29.880	19.920
RACE	41.310	33.780	68.430	81.210
LDSA	-255.600*	58.880	-69.940*	30.530
Constant	-59.852		-28.345	
S.E.	255.199		159.234	
F	(11,388) = 26.020		(10,389) = 13.685	

*Statistically significant at the 0.05 level of the one-tailed t-test.

**Statistically significant at the 0.05 level of the two-tailed t-test.

[a] ABSN refers to the frequency with which the individual missed work for reasons other than illness.

[b] HMPN refers to annual hours of home production, e.g., car repairs, house additions and repairs, etc.

Table 5.13b

Two-Stage Least Squares Estimates of Chronic Illness Expressions (LDSA)

Sample	(1)		(2)	
Variable	β 1970	s	β 1971	s
RISK	-0.030	0.052	-0.016	0.048
AGEH	0.029*	0.006	0.031*	0.005
INSR	-1.475*	0.257	-0.553*	0.170
CHEM	-6.804*	2.479	0.268	0.703
CITY	0.052	0.129	0.050	0.134
POOR	0.500*	0.150	0.345*	0.135
FEDU	-0.036	0.044	-0.028	0.046
ICTR	$-0.17\text{x}10^{-3}$ *	$0.05\text{x}10^{-3}$	$-0.80\text{x}10^{-4}$ *	$0.16\text{x}10^{-4}$
TSPN	0.0021	0.0020	0.0039*	0.0010
SULN	-0.001	0.004	-0.0007	0.0014
WAGE	0.005*	0.002	0.005*	0.001
Constant	0.521		0.033	
S.E.	1.168		1.193	
F	(11,388) = 9.733		(11,388) = 13.250	

*Statistically significant at the 0.05 level of the one-tailed t-test.

**Statistically significant at the 0.05 level of the two-tailed t-test.

coefficient for CHEM in expression (1) is undoubtedly an anamoly since only one person worked in the chemicals and metals manufacturing sector. Those respondents who were poor when growing up demand less chronic illness avoidance, perhaps because their health status is initially less and they therefore must invest more to reach a given health status level. WAGE, and therefore $\widehat{\text{WAGE}}$, is defined here as the individual's marginal hourly earnings. This need not be his hourly earnings without overtime on his primary job.

In Table 5.13b, only expression (2) possesses a statistically significant air pollution coefficient. Assuming as in previous sections that each unit of LDSA is slightly more than two years, or 830 days, in length, that each of these days is a potential workday, and that the average workday is eight hours long, then an additional unit of TSPN in expression (2) of 5.13a causes the individual's utility-maximizing number of days of chronic illness to increase by 3.25 days over his lifetime. We have no idea, however, how these additional days will be distributed over his lifetime, nor can we treat 3.25 additional days for someone who is already chronically ill as similar to 3.25 additional days for someone who is not now chronically ill. Assume our representative individual currently has no chronic illness, and further assume that perpetual exposure to an additional unit of TSPN will cause him to acquire immediately a "chronic" illness. The arithmetic mean for $\widehat{\text{WAGE}}$ in 1971 is \$3.72 per hour. Since our representative individual works eight hours per day, and since he will now find that 3.25 days of his worktime will at some time no longer be available, he would be willing to pay an undiscounted amount of \$96.72 in a single lump sum. The arithmetic mean of TSPN in sample (2) is 156.185, and 127.574 units of TSPN is two standard deviations removed from this mean. Assuming the validity of a linear exprapolation of the preceding marginal (= average) willingness to pay to avoid the chronic illness induced by perpetual exposure to an additional unit of TSPN, we find that the representative respondent would be willing to pay an undiscounted lump sum of approximately \$25,000 (\$24,678) to avoid the chronic illnesses associated with spending the rest of his life in an extremely high TSPN location as opposed to an extremely low TSPN location.

5.7 Overview of Empirical Results

We view the empirical results of this chapter as tentative and ongoing rather than as definitive and final. The SRC interview data that we employ is a random sampling of the civilian, noninstitutionalized population of the United States. Extrapolations of results to the entire population are therefore fairly reasonable, even though we have not employed the SRC sampling weights. However, caution must be exercised in doing so: our measures of illness are substantially less than ideal. In particular, the measure of chronic illness is rather discrete and its uppermost value is unbounded. Moreover, individuals who died between the reference year of the interview and the time of interview are not included. Both factors probably cause the health impact of air pollution to be underestimated. Nevertheless, we feel that we have provided an example of some of the things that can be done with microepidemiological data on health status, endowments, and time and budget allocations. In the bulk of the dose-response expressions we have estimated, most of which were estimated from distinct random samples, air pollution is associated with increased time spent being acutely and/or chronically ill.

Air pollution, in addition, appears to influence labor productivity, where the reduction in productivity is measured in the earnings lost due to reductions in salable skills and in work-time. The reduction in productivity due to air pollution-induced chronic illness seems to overwhelm any reductions due to air pollution-induced acute illness.

The following examples involve linear extrapolations of estimated labor productivity effects and willingnesses-to-pay at arithmetic mean air pollution concentrations. The linear exprapolations extend two standard deviations from the means of the frequency distributions of these concentrations. Geographical locations residing in the upper tails of these distributions might reasonably be regarded as extremely dirty while those along the extended portion of the lower tails are bathed in extremely clean air. The representative individual who is instantaneously and painlessly removed from an extremely dirty location to an extremely clean one might expect to acquire about $20 (in 1970-71 dollars) in additional annual earnings from reductions in air pollution-induced acute illnesses. This same individual would annually acquire several hundred 1970-71 dollars (approximately $100 to $600 in our empirical tests) by the reduction in chronic illness he would obtain from a similar removal. Both these results assume that wage rates are not adjusted in response to a cleaner environment.

The willingness of the representative individual to pay for the annual hours of acute illnesses he could avoid by being in a clean rather than a dirty environment is, for the two samples for which we obtained estimates, between $300 and $500 annually in 1970-71 dollars. For chronic illness avoidance, we calculated, under some extremely crude assumptions and on the basis of only a single sample, that the representative individual would be willing to pay an undiscounted lump sum of $25,000 to be in the clean rather than the dirty environment.

FOOTNOTES

[1] The Survey Research Center possesses the exact addresses of the sample families, but does not include them in its data tapes.

[2] This "errors in variables" problem is usually handled by using instrumental variables which are highly correlated with the variable measured with error but are uncorrelated with the error. We were unable to think of a variable having these properties.

[3] SRC interviews for 1973 behavior and status include a three-digit occupational code corresponding to the coding used in U.S. Bureau of the Census, 1970 Census of Population Alphabetical Index of Industries and Occupations, Washington, D.C.: USGPO (1971). This means that information is available in the SRC data set on nearly the exact kind of job held by the family head and/or his wife. This rather magnificent store of information obviously has many research possibilities which remain completely unexploited in this study.

[4] Other measures of ill-health are available in the SRC data set, particularly the severity of the disability, if any, and the number of weeks missed from work due to sickness. Because of its qualitative nature, the decision was made to use the first of these entirely as an explanatory rather than as a dependent variable.

[5] Expressions of discomfort with the reductionist perspective are now fairly common in the biomedical literature. See, for example, Syme and Berkman (1970) and Engel (1977). More importantly, there is empirical evidence that variations in self-reported health status reflect correct variations in clinically objective measures of health. See Grossman (1975, p. 168) for a review of this literature as well as some additional empirical evidence.

[6] The literature which views children as an investment is surveyed in several papers in a supplement to the March/April 1973 issue of the Journal of Political Economy.

[7] The 18th century French jurist and philosopher, Montesquieu (1947, p. 245), succinctly stated the central theme of much of this literature: "The nations of hot countries are timorous like old men, the nations in colder regions are daring like youngsters." Recent efforts have been considerably less elitist and self-congratulatory.

[8] A fair amount of work appears to have been done to ascertain the discrepancies, if any, between self-reported and clinically evaluated health status. Survey Research Center (1977, pp. 7-10) states that the bulk of studies conclude: (1) as the time between an interview and an event lenghtens, there is increased underreporting about the magintude of the event; (2) important events are less likely to be incompletely and inaccurately reported; and (3) self-reported events are likely to be biased in what the respondent considers to be a socially acceptable direction. Marquis (1978), however, disputes these conclusions because all studies either check self-reported health status against clinical records or check clinical records against status. He shows that a statistically correct test of bias requires both

checks. When he performed this check with a sample of individuals from Dayton, Ohio, he found that "...there is little or no average reporting bias in hospital admission/discharge data obtained by household interviews." (p. 42).

[9] This high proportion of non-whites is probably caused by the fact that 40 percent of the original 1967 sample was composed of families previously interviewed by the U.S. Bureau of the Census for the 1966 Survey of Economic Opportunity.

[10] Wallace (1977) has surveyed a number of recently evolved tests enabling the investigator to ascertain the extent to which Selvin's and Stuart's (1966) data-dredging alter the trustworthiness of later estimates. We have disregarded the Wallace (1977) tests in this study in favor of drawing entirely new samples each time a new expression is estimated.

[11] There is another alternative: each of the following structural expressions could be estimated:

a) DSAB = f(Air pollution, lifestyle, . . ., etc.)

b) LDSA = g($\widehat{\text{DSAB}}$, . . .)

where the $\widehat{\text{DSAB}}$ in expression (b) is the estimated value of DSAB. However, since DSAB is measured in ordinal and discrete terms, either a multinomial logit or a basic logit specification using maximum likelihood estimation methods would have to be employed. In the latter case, four different versions of (a) would have to be estimated since DSAB involves four discrete ordinal measures.

[12] See, for example, Grossman and Benham (1973), Thaler and Rosen (1975), and Parsons (1977).

[13] A review of recent work is available in Lave and Seskin (1977).

[14] In Grossman's (1972) notation, the sick time production function is:

$$TL_t = b_0 + b_1[I_t + \delta)H_{t-1}]$$

where TL represents chronic or acute sick time, H is the stock of health capital, I_t is current health investments, and δ is the rate at which the health stock decays. The term in parentheses is the stock of health capital written in terms of the past stock of health capital and current investment in health. Thus, for example, we treat such variables as POOR, DSAB, and FEDU as determinants of H_{t-1}, and FOOD, NCIG, TSPM, etc., as components of I_t and δ. Grossman (1972) chosses a multiplicative form for this expression, whereas we adopt a linear form. Most inportantly, Grossman (1972) makes both the wage rate and the health state endogenous by making the former a function of the latter and the latter a function of the former. We treat the health state as exogenously determined while retaining the dependence of the wage upon the health state.

[15] This currently lesser productivity could readily be due to past discrimination in the labor market and/or education as well as fewer past opportunities for investment in physical health.

[16] In this case, it is unlikely that multicollinearity has seriously inflated the standard errors of the air pollution variables. The highest simple correlation coefficient between an air pollution variable and another explanatory variable was $r_{TSP \cdot SULM} = 0.65$. All other simple correlation coefficients were less than 0.20.

[17] The 830-day interval is a weighted arithmetic mean established by taking the midpoint of each of the time equivalents of the SRC index measures for LDSA and weighting by the proportion of the entire SRC sample in 1971 having a particular LDSA index value. Ten years was treated as the midpoint for the uppermost LDSA index.

[18] See Crocker and Horst (1977).

[19] Ideal generalization of this model would have: (1) the flow of health services rather than the stock of health entering the individual's utility function; and (2) the opportunity cost of time not be assumed equal to the wage rate but rather derived from the model to be equal to the wage rate weighted by the shadow price for expenditures on inputs into the production of health and the composite commodity.

[20] The time expenditure at which fatugue and/or boredom sets in on a particular job and the rate at which it changes is itself likely to be a function of the individual's state-of-health and education. We have not tried to capture this either in this model or in the subsequent empirical effort that accords with it. One might argue that various attitudinal measures such as job satisfaction, aspiration and ambition, and others readily available in the SRC data would serve as adequate proxies for fatugue and boredom.

[21] Simultaneity is implied by the model in the demand for acute illness. In particular, although we considered it only in passing, the time expended in other work activities is an endogenous variable, which, in turn, implies that work time is endogenous. We have, infact, tested this sumultaneity by treating work as endogenous and estimating the systme by two-stage least squares. The results, which we do not bother to report here, differed only trivially from the ordinary least squares estimates that we do report.

[22] The reader should be aware that by adopting a TSLS estimation procedure, we are giving up some efficiency in estimation in order to enhance the consistency of our estimation. The cruelty of this tradeoff is due to the quite low coefficients of determination involved in OLS estimates of the freedom from chronic illness demand function.

Chapter VI

AN ESTIMATE OF NATIONAL LOSSES IN LABOR PRODUCTIVITY DUE TO AIR POLLUTION-INDUCED MORBIDITY

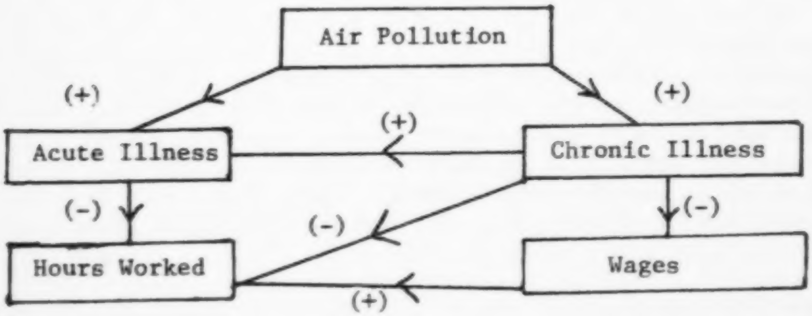
6.1 Introduction

In this brief chapter, we use what we consider to be the most representative of the recursive labor supply estimates in Table 5.10 to speculate what the aggregate gains in U.S. labor productivity could be from a reduction in air pollution-induced acute and chronic morbidity. Due to the preliminary and exploratory nature of our work, we are most anxious that the reader wishing to employ or to communicate these calculations be careful always to make highly visible the set of assumptions in which the calculations are embedded. Otherwise, he will be unable to make an informed judgment about the extent to which the world represented in the text corresponds to reality.

Figure 6.1 is a hueristic representation of the structure forming the basis of our estimate. Air pollution is viewed as increasing directly both chronic and acute illness. In addition, it causes an indirect increase in acute illness via its positive effect on chronic illness. Acute illness reduces hours worked, but, because of its passing nature, it has no impact upon the worker's long-term productivity that determines the level of his wages. However, chronic illness, which does reduce long-term productivity, exerts a direct negative influence on both wages and hours worked. It also influences hours worked in an indirect manner through its effect upon wages.

Figure 6.1

A Representation of the Effect of Air Pollution Upon Labor Productivity



6.2 The Assumptions

Table 6.1 is a succinct list of the major assumptions underlying our empirical implementation of the structure depicted in Figure 6.1 and its extrapolation to a national aggregate. We divide these assumptions into four classes: specification, measurement, estimation, and aggregation. The table also indicates the probable direction of bias, if any, the assumption introduces. However, we do not now know the sensitivity of our estimates and calculations to any particular assumption or to the entire set of assumptions. Upon reviewing Table 6.1, the judicious reader will immediately become aware that our listing is sufficiently strenuous to raise some questions about whether our estimates and calculations are yet sufficiently compelling to warrant their serious use.

In spite of the lengthy listing of assumptions, we emphasize that our treatment of the system in Figure 6.1 has several positive distinguishing features. To balance any negative impressions established from Table 6.1, we list these positive features in Table 6.2. Our estimates of the system in Figure 6.1 is presented in Table 6.3. As a result of a one-unit ($\mu g/m^3$) increase in air pollution, we estimate that the representative person in Table 6.3 will have his annual work hours reduced by 0.547 hours. Of this reduction, only 0.046 hours will be due to acute illness. The loss in labor productivity suffered by this person can be calculated by (where Δ stands for change):

$$\frac{\Delta(\text{Work hours} \cdot \text{Wage})}{\Delta(\text{pollution})} = \frac{\Delta(\text{Work hours})}{\Delta(\text{Pollution})} \cdot \text{Wage} + \frac{\Delta(\text{Wage})}{\Delta(\text{Pollution})} \cdot \text{Work hours}$$

Upon performing this calculation, we obtain:

$$= (0.547)(\$3.225) + (\$0.071)(1560.895)$$

$$= \$2.86$$

That is, a one-unit reduction in air pollution would have increased this representative person's 1970 earnings by $2.86. Only $0.15 of this sum represents the gain from a reduction in acute illness.

The above $2.86 sum represents our "best" estimate at this point of the representative person's gain in 1970 earnings from a one-unit reduction in air pollution. Lower and upper bounds for this estimate can be established by making use of the confidence intervals for the effect of pollution on chronic and acute illness; that is, we wish to calculate the gain in earnings when the pollution coefficient in (1) is 0.0028 ± 0.0011, and when the pollution coefficient in (2) is 0.623 ± 0.317. At least for the chronic illness expression, this confidence interval captures nearly all the range of the values for the pollution coefficients in the chronic illness expressions estimated in the previous chapter. Upon performing this calculation for the lower bound, we obtain $1.88, and for the upper bound, we obtain $3.84.

Assume that the average exposure of the U.S. 1970 urban population to annual geometric mean total suspended particulates was 100 $\mu g/m^3$ and that

Table 6.1

Major Assumptions Limiting Generality of Results

Specification

1) Air pollution affects only the duration of chronic illness. Our inattention to the severity of chronic illness tends to reduce the estimated impact of air pollution on labor productivity.

2) Occupational exposures to hazards and environmental pollutants other than air pollution do not influence either acute or chronic illness. If air pollution is moderately and positively associated with these hazards and pollutants, this assuption tends to increase the estimated impact of air pollution on labor productivity.

3) Annual geometric mean ambient concentrations of total suspended particulates serve as an adequate proxy for all forms of air pollution. The effect of this assumption upon the estimated effect of air pollution on labor productivity is unknown.

4) All relationships depicted in Figure 6.1 are linear. It is unknown what effect this assumption has on the estimated effect of air pollution on labor productivity.

5) Air pollution-induced health effects do not cause the voluntary substitution of leisure for work. This assumption tends to reduce the estimated impact of air pollution on labor productivity.

Measurement

6) Air pollution exposures for each individual in the sample are adequately represented by a single annual average of ambient concentrations obtained at a single monitoring station within the individual's county of residence. Since pollution monitoring stations in the early part of the 1970's were predominantly in downtown urban locations, individuals' air pollution exposures probably tend to be exaggerated. This will reduce the estimated health effects of air pollution.

7) The duration of any air pollution-induced chronic illness cannot exceed ten years. This will reduce the estimated effect of air pollution upon the duration of chronic illness.

8) Housewives, retirees, and students, who together constitute about twenty percent of our samples, do not contract air pollution-induced acute illnesses. This assumption will tend to reduce the estimated impact of air pollution upon labor productivity.

9) Air pollution-induced chronic and acute illnesses are a constant proportion of all illnesses. The effect of this assumption is unknown.

(continued)

Table 6.1

(continued)

10) The quantity of preventive and ameliorative medical care an individual consumes is adequately measured by whether or not he has medical insurance. This assumption has an unknown effect upon our estimates.

11) Relative air pollution concentrations across the U.S. have been fairly constant. This assumption has an unknown effect upon our estimates of air pollution-induced chronic illness.

12) When interviewed, the individuals in the sample had no incentive to bias their answers nor did they have difficulty accurately recalling their personal medical histories of the previous twelve to sixteen months. The effect of this assumption upon our estimates is unknown.

13) No individual who would otherwise have been included in the sample died between the time for which information was to be gathered and the time of the interview. In fact, about five percent of the potential respondents died each year. The effect of this assumption is to reduce the effects of air pollution upon labor productivity.

Estimation

14) With the available data, classical linear regression procedures provide consistent and unbiased estimates of the structure depicted in Figure 6.1. The effect of this assumption upon our estimates is unknown.

Aggregation

15) The response of the health state of each individual in the U.S. to any given change in ambient air pollution is a constant. The effect of this assumption upon the calculation for the aggregate effect of air pollution upon labor productivity is unknown.

16) The response of the health state of every individual in the U.S. to ambient air pollution changes is identical. The effect of this assumption upon the calculation for the aggregate effect of air pollution upon labor productivity is unknown.

Table 6.2

Distinguishing Features that Enhance the Generality of Results

1) The acute illness and chronic illness dose-response estimates used to calculate the aggregate impact of air pollution-induced morbidity upon U.S. labor productivity are representative of estimates obtained from many different independent samples drawn from the same data set. In effect, substantial quasi-replication of the dose-response estimates has been performed.

2) The system is estimated only for people who have always lived in one state. We believe this restriction enhances the extent to which we capture the effect of the history of air pollution exposures upon the chronic illness dose-response function.

3) Our estimated expressions for wages and hours worked are very similar to those obtained by other economists.

4) We include more information on life-styles and genetic and social endowments than is usually included in dose-response expressions estimated from epidemiological data.

Table 6.3

Estimated Expressions to be Used to Calculate the Effect of Air Pollution-Induced Illness on Labor Productivity[a]

(1) 830 day years chronically ill = 2.980 + 0.554 (illness severity)** +
(0.035)
0.005(age in years) + 0.013(years of school) - 0.044(father's years of
(0.004) (0.029) (0.037)
school) - 0.069(poor when growing up) + 0.072(Caucasoid) + 0.139(male)
(0.103) (0.488) (0.114)
- 0.902(diet adequacy) - 0.454(has medical insurance)* - 1.645(works
(0.975) (0.129) (0.575)
in chemicals/metals industries)[b] + 0.0028(mean total suspended
(0.0011)
particulates)*

$R^2 = 0.525$; S.E. = 0.964; F(11,388) = 38.920.

(2) Annual hours acutely ill = 165,208 + 39.52(years chronically ill)*
(13.34)
-1.421(age in years) - 16.92(male) - 0.086(cigarette expenditures) -
(1.312) (39.16) (0.118)
78.47(gets strenuous exercise)* - 0.105(diet adequacy)* - 38.44(degree
(40.11) (0.033) (13.26)
of risk aversion)* + 187.70(has medical insurance)** - 85.56(works in
(47.47) (191.20)
chemicals/metals industries) + 0.623(mean total suspended particulates)*
(0.317)

$R^2 = 0.195$; S.E. = 204.462; F(10,389) = 5.721.

(3) Wage in cents = -132.318 - 25.930(years chronically ill)* + 24.070(years
(14.440) (8.578)
of school)* + 15.370(illness severity) + 26.880(family size)* + 42.380
(18.260) (6.079) (6.138)
(cost-of-living)* = 52.950(years on current job)* - 7.163(often late
(22.130) (33.88)
for work) + 66.090(union member)* + 47.60(Caucasiod)
(34.580) (34.22)

$R^2 = 0.408$; S.E. = 258.908; F(11,388) = 24.28.

(continued)

Table 6.3

(continued)

$$\text{(4) Annual hours worked} = 1266.68 - \underset{(27.22)}{163.90}\text{(years chronically ill)}^{*} + \underset{(0.130)}{0.354}\text{(wage in cents)}^{*} + \underset{(16.68)}{44.26}\text{(family size)}^{*} + \underset{(80.27)}{519.80}\text{(male)}^{*} - \underset{(0.022)}{0.272}\text{(dollars Of transfer income)}^{*} + \underset{(15.20)}{23.06}\text{(cost-of-living)} - \underset{(0.031)}{0.074}\text{(annual hours acutely ill)}^{*}$$

$R^2 = 0.551$; S.E. = 663.196; $F(6,393) = 80.41$.

[a] Exact variable definitions are available in Table 5.1.

[b] The number of people in these industries was too small for this coefficient to be meaningful.

*Statistically significant at the 0.05 level of the one-tailed t-test.

**Statistically significant at the 0.05 level of the two-tailed t-test.

the standard deviation of these exposures was 30 $\mu g/m^3$. Throughout this study, total suspended particulate measures have been highly correlated with other air pollutants so that total suspended particulates probably serve as an adequate proxy for all air pollution. Further assume that the national urban population is approximately 150 x 10^6 people, each of whom is or will be a family head. After age 20, each of these family heads has a life-span of 50 years and any air pollution-induced chronic illnesses he contracts are distributed rectangularly over the 50 years. The earnings he loses due to the presence of an acute or chronic illness do not vary over the years. Given these and earlier assumptions, a 60 percent reduction in air pollution would, in June 1978 dollars, increase the value of 1970 U.S. labor productivity by the amounts shown in Table 6.4. Most of the gain would accrue due to reductions in air pollution-induced chronic illness.

It must be strongly emphasized that the magnitudes exhibited in Table 6.4 are extremely sensitive to the assumptions we have made. Nevertheless, given any reasonable set of assumptions about air pollution exposures, size of the population exposed, etc., the estimates of labor productivity gains in Table 6.4 are much larger than previous estimates of all types of annual gains from air pollution control in the United States. No gains in labor productivity, via reductions in air pollution-induced health effects, have previously been developed. It thus appears that the economic gains from the morbidity effects of air pollution control have been greatly undervalued, perhaps because most prior research efforts have concentrated upon mortality rather than morbidity.

A more conservative but equally tenuous way of calculating the effects in Table 6.4 might proceed as follows. Assume that the 75 percent, or 112 x 10^6 million people of the 150 x 10^6 urban population are 16 years or older. At age 16, each of these adults has a lifespan of 56 years and any air pollution-induced chronic illnesses he contracts are distributed rectangularly over the 56 years. The annual earnings he loses due to the presence of an acute or chronic illness do not vary over the 56 years. If the median household size is 2.0, there are then 56.25 x 10^6 urban household heads. There is thus a \$160.88 x 10^6 = (\$2.86) (56.25 x 10^6) gain in the labor productivity for household heads from a one unit reduction in air pollution.

If two-thirds of the household heads are married, if 35 percent of these households have working wives, and if working wives earn 60 percent as much as their male counterparts, there would then be a \$22.58 x 10^6 = (\$2.86) (0.6) (13.13 x 10^6) gain in the labor productivity of working wives.

If the value of household services provided by all household members in each urban household is 40 percent of the household head, there would then be a \$64.35 x 10^6 = (\$2.86)(0.4) (56.25 x 10^6) gain in the household labor productivity of all urban households. Adding the results for household heads, working wives, and household labor, we obtain a \$247.81 x 10^6 gain in labor productivity for a one unit reduction in air pollution. A 60 percent reduction in 1970 air pollution would then, in August 1978 dollars, increase the value of 1970 urban labor productivity by \$25 x 10^9 dollars. This is a "best" estimate. Its upper and lower bounds are, respectively, \$34 x 10^9 and \$16 x 10^9. If one performs these identical calculations in precisely the same fashion for a 1977 U.S. total population of 216.1 x 10^6, one obtains a "best" estimate of \$36 x 10^9.

Table 6.4

Estimated Per Capita Aggregate Gains in 1970 U.S. Labor Productivity Due to a 60 Percent Reduction in Air Pollution

(June 1978 Dollars)

	Per Capita	Aggregate
Lower Bound	\$189.50	$28,426 \times 10^6$
"Best" Estimate	\$288.29	$43,243 \times 10^6$
Upper Bound	\$387.07	$58,061 \times 10^6$

REFERENCES

Becker, G.S., "A Theory of the Allocation of Time," Economic Journal 75 (September 1965), 493-517.

Blumquist, G., "Value of Life: Implications of Automobile Seat Belt Use," Unpublished, March 1977.

Conley, B., "The Value of Human Life in the Demand for Human Safety," American Economic Review 66, (March 1976), 45-57.

Crocker, T.D. and Horst, R.L, Jr., Oxidant Air Pollution and Work Performance of Citrus Harvest Labor, Research Triangle Park, N.C.: USEPA Publication, 600/5-77-013 (September 1977).

Cropper, M.L., "Health, Investment in Health, and Occupational Choice," Journal of Political Economy 85 (December 1977), 1273-1294.

Engel, G.L., "The Need for a New Medical Model: A Challenge for Biomedicine," Science, (1977), 129-136.

Grossman, M., "On the Concept of Health Capital and the Demand for Health," Journal of Political Economy 80 (March/April 1972), 223-255.

Grossman, M., and Benham, L., "Health, Hours, and Wages," in M. Perlman, (ed.), The Economics of Health and Medical Care, New York: Halstead Press (1973), 205-233.

Grossman, M., "The Correlation between Health and Schooling," in N.E. Terleckyj (ed.), Household Production and Consumption, New York: Columbia University Press (1975), 147-211.

Hanushek, E.A., and Quigley, J.M., "Implicit Investment Profiles and Intertemporal Adjustments of Relative Wages," American Economic Review 68 (March 1978), 67-79.

Hippocrates, 460-377 B.C., "On Airs, Waters, and Places," in F. Adams, (ed.), The Genuine Works of Hippocrates, Baltimore (1939).

Kneese, A. V., and Schulze, W., "Environment, Health and Economics - The Case of Cancer," American Economic Review 67, (February 1977), 26-32.

Kmenta, J., Elements of Econometrics, New York: Macmillan Publishing Company (1971).

Lave, L. and Seskin, E., Air Pollution and Human Health, Baltimore: Johns Hopkins Press (1977)

Lazear, E., "Age, Experience, and Wage Growth," The American Economic Review 66 (September 1976), 548-558.

Marquis, K. H., Record Check Validity of Survey Responses: A Reassessment of Biases in Reports of Hospitalizations, Report R-2319-HEW, Santa Monica, California: The Rand Corp. (May 1978).

McCarthy, M.D., "Notes on the Selection of Instruments for Two Stage Least Squares and K Class Type Estimators of Large Models," Southern Economic Journal, (January 1971), 251-259.

Mincer, J., "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach," Journal of Economic Literature 8 (March 1970), 1-26.

Mishan, E.J., "Evaluation of Life and Limb: A Theoretical Approach," Journal of Political Economy, Vol. 79, (December 1971).

National Academy of Science, Nitrates: An Environmental Assessment, Environmental Studies Board, NAS, Washington, D.C., 1978.

Schwing, R.C. and McDonald, G.C., "Measures of Association of Some Air Pollutants, Natural Ionizing Radiation and Cigarette Smoking with Mortality Rates," presented to the International Symposium on Recent Advances in the Assessment of the Health Effects of Environmental Pollution, Paris, France (June 1974).

Schultz, T.P., The Estimation of Labor Supply Functions for Secondary Workers, Report R-1265-NIH/EDA, Santa Monica, California: The Rand Corporation (February 1975).

Selvin, H.C. and A. Stuart, "Data-Dredging Procedures in Survey Analysis," The American Statistician 20 (1966), 20-23.

Smith, R., "The Feasibility of an 'Injury Tax' Approach to Occupational Safety," Law and Contemporary Problems, (Summer-Autumn, 1974).

Stigler, G.J., "Imperfections in the Capital Markets," Journal of Political Economy 75 (June 1967) 291.

Survey Research Center, University of Michigan, Experiments in Interviewing Techniques: Field Experiments in Health Reporting, 1971-1977, Washington, D.C." USDHEW Publication No. (HRA) 78-3204 (Nov. 1977).

Survey Research Center, A Panel Study of Income Dynamics: Study Design, Procedures, Available Data, Ann Arbor: Institute for Social Research, University of Michigan (1972a).

Syme, S.L. and Berkman, L.E., "Social Class, Susceptibility and Sickness," American Journal of Epidemiology 104 (1976), 1-8.

Thaler, R. and Rosen, S., "The Value of Saving a Life: Evidence from the Labor Market," in N.E. Terleckyj, (ed.), Household Production and Consumption, New York: Columbia University Press (1975), 265-297.

Waddell, T.E., The Economic Damages of Air Pollution, Washington, D.C.: Socioeconomic Environmental Studies Series, USEPA (1974).

Wallace, T.D., "Present Estimation in Regression: A Survey," American Journal of Agricultural Economics 59 (August 1977), 431-443.